

AD-A201 678

## Reply to the Discussion of "Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Resource"

*John Haslett*

Department of Statistics  
Trinity College  
Dublin 2  
Ireland.

*Adrian E. Raftery*

Department of Statistics, GN-22  
University of Washington  
Seattle, WA 98195  
USA.

DTIC  
ELECTED  
NOV 25 1988  
S D

### ABSTRACT

The paper "Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource" (Technical Report No. 110, Department of Statistics, University of Washington) was read before the Royal Statistical Society at a meeting organized by the Research Section on May 25, 1988. There were 33 discussants, who between them made more than 100 separate suggestions and queries. This is the reply to the Discussion; the contributions to the Discussion are included as an Appendix.

Many of the discussants were concerned that the model used was not sufficiently general. We argue that this is not a problem for the present application, but we do propose a more general model for use in other contexts. This allows for non-homogeneity of temporal dependence across sites and for anisotropic spatial correlation. We review the evidence for long-memory dependence as opposed to non-stationarity, and for the use of fractional differencing to model it. We discuss computational and asymptotic aspects of the estimation of the fractional differencing parameter, the location parameter of the wind speed distribution, and the distribution of wind power. Many other points are discussed, including the order in which transformation and aggregation are carried out and the treatment of the "outlier" Rosslare.

---

John Haslett is Senior Lecturer, Department of Statistics, Trinity College, Dublin 2, Ireland. Adrian E. Raftery is Associate Professor of Statistics and Sociology, Department of Statistics, GN-22, University of Washington, Seattle, WA 98195, USA. Raftery's research was partially supported by ONR contracts N00014-84-C-0169 and N00014-81-K-0095. The authors are grateful to Julian Besag, Liam Burke, Michael Newton, Paul Sampson and Richard Smith for helpful discussions during the preparation of this reply.

DISTRIBUTION STATEMENT A  
Approved for public release  
Distribution Unlimited

88 11 14 065

**Reply to the Discussion of "Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Resource"**

*John Haslett*

Department of Statistics  
Trinity College  
Dublin 2  
Ireland.

*Adrian E. Raftery*

Department of Statistics, GN-22  
University of Washington  
Seattle, WA 98195  
USA.

DTIC  
S ELECTED NOV 25 1988 D

**ABSTRACT**

The paper "Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource" (Technical Report No. 110, Department of Statistics, University of Washington) was read before the Royal Statistical Society at a meeting organized by the Research Section on May 25, 1988. There were 33 discussants, who between them made more than 100 separate suggestions and queries. This is the reply to the Discussion; the contributions to the Discussion are included as an Appendix.

Many of the discussants were concerned that the model used was not sufficiently general. We argue that this is not a problem for the present application, but we do propose a more general model for use in other contexts. This allows for non-homogeneity of temporal dependence across sites and for anisotropic spatial correlation. We review the evidence for long-memory dependence as opposed to non-stationarity, and for the use of fractional differencing to model it. We discuss computational and asymptotic aspects of the estimation of the fractional differencing parameter, the location parameter of the wind speed distribution, and the distribution of wind power. Many other points are discussed, including the order in which transformation and aggregation are carried out and the treatment of the "outlier" Rosslare.

---

John Haslett is Senior Lecturer, Department of Statistics, Trinity College, Dublin 2, Ireland. Adrian E. Raftery is Associate Professor of Statistics and Sociology, Department of Statistics, GN-22, University of Washington, Seattle, WA 98195, USA. Raftery's research was partially supported by ONR contracts N00014-84-C-0169 and N00014-81-K-0095. The authors are grateful to Julian Besag, Liam Burke, Michael Newton, Paul Sampson and Richard Smith for helpful discussions during the preparation of this reply.

We would like to thank the many discussants for their kind and penetrating contributions. We are grateful to all for making their remarks so relevant to the paper! We apologise if our reply overlooks some of the 100 or so separate suggestions and queries.

Our project had a specific goal, namely the estimation of the mean kinetic energy in the wind at a site for which only a short run of data is available. To do this, we produced a model which was easy to apply at a new site, exploiting the remarkable empirical regularities highlighted by Dr. Carlin and Dr. Ray. We could have developed a more complicated model which might have better described some fairly minor features of the synoptic data, but this would have made the method harder to apply at a new site, and numerical work referred to in Section 6 indicates that it would not have improved the results. Modelling the existing data was not an end in itself.

Nevertheless, Professor Smith rightly says that the wide range of potential applications justifies looking for models more general than (4.1). Indeed, more than half the discussants suggested ways of elaborating the model. Equation (4.1) is a special case of the general model

$$\Phi(B)(Z_t - \mu - s_t) = \nabla^{-d} \Theta(B) \varepsilon_t. \quad (A)$$

In (A),  $Z_t$  is the vector of undeseasonalized velocity measures on day  $t$ ,  $\Phi(B) = I - \Phi_1 B - \dots - \Phi_p B^p$ , and  $\Theta(B) = I - \Theta_1 B - \dots - \Theta_q B^q$ , where  $\Phi_1, \dots, \Phi_p$  and  $\Theta_1, \dots, \Theta_q$  are  $m \times m$  matrices such that the zeros of the determinantal polynomials  $|\Phi(B)|$  and  $|\Theta(B)|$  are outside the unit circle,  $\mu = (\mu_1, \dots, \mu_m)^T$ ,  $s_t = (s_{1t}, \dots, s_{mt})^T$  is a vector of seasonal effects,  $\nabla^{-d} = (\nabla^{-d_1}, \dots, \nabla^{-d_m})^T$ , and  $\varepsilon_t \stackrel{\text{iid}}{\sim} \text{MVN}(0, V)$ . As Dr. McLeod points out, (A) is an extension and synthesis of many proposals in the literature, most of which are cited in Camacho, McLeod and Hipel (1987a).

If (A) is unconstrained, parameters proliferate wildly, as Professor Dempster has noted. Each parameter in (A) is associated with either a single site or a pair of sites, and so may be constrained to be a function of position and/or (directed) separation which is either (1) constant; (2) deterministic and parametric; (3) deterministic and non-parametric; or (4) stochastic. Our

model (4.1) is based on constraints of types (1) and (2), while several discussants suggest constraints of type (3). Stochastic constraints lead to parametric empirical Bayes models (Deely and Lindley, 1981; Morris, 1983). This is intellectually the most satisfying approach, but it is also the most difficult, and only Professor Ogata has had the courage to tackle it.

Model (A) encompasses virtually all the suggestions for model elaboration made by discussants. With suitable adaptation, the methods of statistical analysis developed in Section 4 may be applied to it.

#### *Data analysis*

Dr. Kent's comparison of the square root transformation at different levels of aggregation with the log normal transformation elsewhere is perceptive; Carlin and Haslett (1982) found this effective for hourly data. He is right in his surmise that transforming and aggregating could have been performed in reverse order. This might indeed have led to a simpler approach than in Section 5, as implicitly sought by a number of contributors concerned with power considerations. It may be of interest, however, that one of the practical criticisms levelled at our solution by our meteorological colleagues is that our method, developed for data disaggregated to the level of days, is applicable with difficulty to a number of valuable short runs of data already available, but published solely as means, and to data that might be collected by particularly cheap 'run-of-the-wind' anemometers which simply return a mean wind speed for the observation period. Such data cannot be disaggregated to days, never mind hours, before transformation. Our general approach can be used for such data, but the details of the method require modification.

Dr. Kent's components of wind-speed model has been used in the literature (McWilliams, Newman and Sprevak, 1979) for hourly data. Almost uniformly preferred is the Weibull model, and Carlin and Haslett's (1982) square root transformation is related to a classical transformation of Wiebull data to normality (Dubey, 1967; Johnson and Kotz, 1970).

Professors Guttorp and Sampson ask whether seasonal variation could be modelled using meteorological theory. We know of no way of doing this. Wind arises because of temperature differences, so the (relatively weak) seasonal pattern in wind speeds is related to a superposition of (usually much stronger) temperature patterns at different places. This, together with atypical wind patterns around the equinoxes, may suggest a meteorological explanation for the need to use several harmonics which troubled Dr. Ray.

For simplicity and ease of application at a new site, we assumed the seasonal effect to be constant throughout Ireland, although, as Professor Ogata points out, there are slight differences between stations. His proposals for modelling these differences are interesting, and we hope that he will try them out on our data.

#### *Rosslare*

Rosslare is an outlier because the correlations with the other stations are too low. We simply removed it from the analysis. Professor Switzer points out that if there are potential sites of interest nearby, this could be an important waste of data. In Ireland, the main sites of interest for wind energy are in the west and the northwest, so that the removal of Rosslare in the southeast is not a problem.

Of course, if the outlying station had been in a location of interest for wind energy, we could not have dealt with it so simply. Professor Lewis proposes an excellent practical way of overcoming the difficulty which, combined with Professor Titterington and Mr. Jamieson's suggestion of a change in  $\beta$ , suggests a whole battery of ad-hoc ways of dealing with isolated particularities in spatial covariance structures. Professors Guttorp and Sampson and Professor Switzer outline more general methodologies for dealing with non-stationarities in the spatial covariance structure, on which we comment later.

Dr. Jolliffe speculates that the unusual behaviour of Rosslare may be due to local topography rather than to a regional effect. The meteorologists, frankly, are puzzled. The station is sited somewhat unfortunately in that the winds from the prevailing direction tend, rather more

than should be the case in ideal circumstances, to pass over the village. But departures from the ideal siting can apparently be found at all stations.

He also says that the lower cross-correlations between Rosslare and other stations could be a by-product of lower autocorrelations at Rosslare. We find it hard to see dramatic differences between the autocorrelations at Rosslare and other stations from Fig. 4 and Fig. 5; in particular, the pattern at Rosslare is similar to those at Roche's Point and Valentia. Roche's Point provides an informal test of Dr. Jolliffe's hypothesis. The cross-correlation between Rosslare and Roche's Point is about one-quarter less than would be predicted from (3.3). Inspection of Fig. 4 indicates that the short-term autocorrelation structures at both stations are well approximated by AR(1) models, while Fig. 5 shows that the long-range dependence patterns are also similar. Dr. Jolliffe's own calcuation yields  $K = 0.9994$ , so that differences in autocorrelations are unlikely to explain the difference in cross-correlation.

Dr. Jolliffe also asks whether we extended the cross-validation exercise to predict the values for Rosslare. Some cross validation on Rosslare was indeed performed. Using 52 weeks of data at Rosslare, a 5 year mean wind was predicted by  $\hat{\mu}_k$  with an error of 1.0%; this error ranked 5th smallest of the 12. For a longer 18 year mean the error was 1.5% which ranked 2 out of 12. This is perhaps another example of the remarkable ( $p < .10$ ) good fortune pointed out by Dr. Glaseby!

#### *Spatial covariance structure*

To respond briefly to Drs. Chatfield and Yar, kriging can indeed be viewed as a minimum mean square error interpolator or predictor in a stochastic process context. Cressie (1985) reminds us that it has been re-invented many times and is similar, for example, to the well known Wiener Filter. Professor Mardia's discussion shows yet another familiar face of the technique. It is more frequently applied in spatial problems with no time replication. A key step in kriging is the estimation of the relationship between (spatial) correlation and distance. In our case, as Professor Tong points out, we model the (temporal) cross-correlation of the  $\epsilon_{it}$ 's as a

function of distance. In this sense the method can be thought of as a multiple time series model, as Professor Stein remarks.

We must disappoint Professor Titterington and Mr. Jameison: we have declined to interpolate the mean wind speed from other means, for with only 12 data points and the expectation on physical grounds of spatial non-stationarity of the mean, this would be foolhardy. Nevertheless the similarities between the difficulties arising in the two problems are important, and many contributors have drawn attention to the fact that we have available here (as typically in geostatistics), very little evidence to guide us at short range separation. As Professors Cressie and Pesarin point out, we did have some additional data. This could, indeed, have been used to adjudicate between the suggestions by Professor Conradsen, Professor Stein and others that the nugget effect is greater than we estimated, and that of Dr. Li that it be ignored altogether. In retrospect these data, which were omitted on meteorological advice due to length of record in one case and anomalous data in the other, could have proved useful here. As Professor Conradsen points out, a change in the variogram structure can have dramatic effects on the kriging weights. What is at issue here, however, is the variance of the difference between an optimal and a sub-optimal estimator, based on a correct and an incorrect variogram, respectively. This is not as dramatic; see comment (8) by Professors Cressie and Pessarin. Of course the correct estimation of the 'kriging variance' does depend critically on the variogram, as remarked by Professor Mardia.

Professor Smith, Professor Lewis, Professors Guttorp and Sampson and Professor Switzer are all concerned that (3.3) is not general enough. Our numerical work indicated that, for our purpose, precision is not greatly improved even by assuming knowledge of the exact spatial covariance structure at the new site, which is presumably the best one can do. Thus (3.3) appears to be general enough for our application. However, in view of other potential applications, it does seem worth considering generalizations, especially as doing so is unlikely to complicate the statistical analysis greatly.

One such is suggested by Professor Smith, who asks whether (3.3) could not be generalized to allow for directional dependence. This is not too difficult. If  $\phi_{ij}$  is the angle of the line joining stations  $i$  and  $j$ , restricted to the range  $[\phi_0, \phi_0 + \pi]$  for some  $\phi_0$ , then one can replace the lower equation in (3.3) by

$$r_{ij} = \exp[-g(d_{ij}, \phi_{ij})]. \quad (B)$$

A simplification of (B) which may often be reasonable is to set

$$g(d, \phi) = g_1(d) + g_2(\phi). \quad (C)$$

In (C) one may specify functional forms such as  $g_1(d) = \alpha + \beta d$ , and  $g_2(\phi) = \exp[\kappa \{ \cos(\phi - \bar{\phi}) - 1 \}]$ , suggested by the von Mises distribution. This could represent a situation in which correlation is strongest along a direction  $\bar{\phi}$ , and declines as one deviates from that direction. For the wind data, however, generalizations such as (B) do not seem necessary.

Professor Lewis finds the lack of directional information counterintuitive. We were disappointed also. To give flavour to this, some observed correlations are shown in Table D1. For simplicity we confine attention to Belmullet, and its correlation with other stations, in 1970.

TABLE D1  
*Correlations between wind speeds at Belmullet, and other stations 1970, based on (transformed) daily averages, and daily averages of (signed) E-W, N-S components.*

	Rpt	Val	Ros	Kil	Sha	Bir	Dub	Cla	Mul	Clo	Mal
Sq. root	.57	.70	.35	.65	.75	.78	.70	.87	.77	.80	.89
E-W	.02	.33	.16	.31	.15	.00	.13	.29	.08	.34	.39
N-S	.21	.29	.18	.09	.12	-.04	-.06	.30	.06	.25	.23

We speculate that aggregation is the source of the difficulty, and that more detailed modelling at the level of hours would be needed to properly exploit this directional information. This would probably need greater attention to be paid to lagged correlations reflecting the weather systems, as suggested by Dr. Henstridge, and, if we understand Professor Mardia's final

point correctly, to cross-covariances between components. We feel that this would contribute little extra at the end of the day.

Professors Guttorp and Sampson outline a non-parametric method for estimating non-stationary and anisotropic spatial covariances. This looks promising, and reveals subtle but potentially important features of the wind data which could not easily be detected otherwise. It also accommodates Rosslare in a smooth way, and provides estimates of the spatial covariance at all locations. A remaining question is whether the estimated covariance structure is guaranteed to be positive definite.

Professor Switzer's alternative proposal is interesting because it provides a way of modifying the assumed global spatial covariance to take account of local structure. However, it is designed for the situation where no data is available at the new site, which was not the case for us. Also, it is not guaranteed to yield a positive definite spatial covariance matrix. Ideally, such a proposal should give weight to the data at a new site that increases with its amount. Devising a scheme which weights data at a new site appropriately while preserving positive definiteness seems to be a real challenge.

Dr. Taam and Professor Yandell suggest setting the problem in a Bayesian context of multivariate smoothing splines. This is an interesting idea, although the problems of implementation seem formidable, and we look forward to more research on this topic. Their more specific proposals for the situation where the data are on a lattice are also interesting, although they do not seem directly relevant to the present problem.

### *Why long-memory?*

Meteorologists have long been aware that the sample mean may exhibit behaviour inconsistent with short-memory dependence, which they often call "potential predictability" (Madden, 1976; Shukla and Gutzler, 1983; Trenberth, 1985). However, as Dr. Glasbey and Dr. Katz point out, they have tended to attribute such behaviour to the rather vaguely defined concept of "climatic drift", which they clearly think of as a form of non-stationarity. By contrast, in the closely related area of hydrology, similar phenomena are often observed, and long-memory dependence is widely accepted as an explanation for them.

We continue to believe that wind speeds in Ireland probably do exhibit long-memory dependence. The decrease in the empirical MSE's in Table 1 seems too rapid to be compatible with most reasonable models for non-stationarity in the mean. Further, certain kinds of behaviour often described as "climatic drift" can be represented by long-memory processes. Dr. Glasbey reports the meteorologists' rule-of-thumb that climatic drift manifests itself in periods greater than 30 years. For a fractionally-differenced model with our estimated  $d = 0.328$ , the variance of a 30-year mean is about the same as that of the mean of 25 independent daily observations! Thus our model implies that disjoint 30-year periods may have quite different means, giving the appearance of climatic drift.

Professor Dempster points out that Fig. 5 does not conclusively establish that the data have a long-memory component, rather than, say, cycles of lengths close to the 11 and 22 year sunspot cycles. In support of the long-memory hypothesis, we can only point to the empirical behaviour of the sample means in Table 1, the lack of apparent cycles or monotonic trends in plots of long series of annual means (up to 40 years) such as those in Raftery *et al.* (1982), and the analogy with hydrology. Professor Dempster also says that the AR(9) filter is capable of representing something indistinguishable from long-memory dependence via roots near unity. However, an autoregressive root near unity cannot account for behaviour of the kind we observed, such as the behaviour of the sample means, which is characteristic of long-memory dependence, but quite different from non-stationarity.

Drs. Chatfield and Yar ask whether some of the long-memory dependence could be explained by the imperfect nature of the seasonal filter, also pointed out by Professor Ogata. Fig. 5 shows that this cannot be so. At each station there is a local peak in the periodogram around the annual frequency resulting from the failure to remove all the seasonal variation, but this is well separated from the low frequency ordinates which reveal the long-memory dependence.

Dr. Henstridge suggests that some of the long-memory effect may be due to changes in measuring equipment and in the environment around the stations, and perhaps even to displacements of the stations themselves. Apparently the measuring equipment has not been changed, except in respect of Malin Head, where the anemometer was raised about 1965; an empirical adjustment (similar to that suggested by Mr. Brontë-Hearne) was made here to preserve continuity. Urban spread has latterly reached some of the stations, originally placed 2-3 miles from the towns. But it seems that during the period 1961-78, this was not regarded as a problem.

#### *Why fractional differencing?*

Several discussants suggested ways of modelling the observed long-term dependence other than fractional differencing. Drs. Chatfield and Yar wonder why we did not use first differencing. The reason is that this yields a non-stationary model of random walk type, which would conflict with the behaviour of the sample means in Table 1.

Dr. Jones suggested a medium-memory model. This is interesting, although the three-parameter model written down is formally a short-memory one, and the behaviour of the sample mean would reflect this. Thus it seems unlikely that such a model could adequately account for the empirical MSE's in Table 1. However, the idea of defining the model in terms of the partial autocorrelations is valuable; in this connection we would draw attention to the pioneering paper of Ramsey (1974), which is often overlooked.

Professor Tong suggests another model, but he is not sure whether or not it has the long-memory property. It is appealingly simple, and so we hope that he, Professor Künsch and Dr. Tjøstheim continue this research.

Dr. Beran points out that long-range dependence may exist in space as well as in time, and Dr. Renshaw has made a real start on modelling it.

#### *Model elaboration*

Professor Smith, Drs. Chatfield and Yar, Professors Cressie and Pesarin and Dr. Li are concerned that forcing the ARMA coefficients to be constant across sites in (4.1) may be unduly restrictive, while Dr. Jolliffe, Professor Tong and (implicitly) Dr. Henstridge suggest allowing direct dependence of  $X_{it}$  on  $X_j^{t-1}$  for  $j \neq i$ . Professors Guttorp and Sampson suggest allowing a gradient in variance across Ireland. All these suggestions lead to special cases of model (A). We chose (4.1) after experimenting with other special cases of (A) because it was the simplest model which enabled us to achieve our objective, not because it captures every feature of the synoptic data.

Based on Fig. 4, Professors Cressie and Pesarin comment that Valentia, Roche's Point and Rosslare do not seem to have the same long-range dependence as the other stations. Detailed features of empirical autocorrelation functions such as those in Fig. 4 are notoriously difficult to interpret, and we preferred to rely on Fig. 5 which indicates that the low-frequency characteristics at these three stations are actually similar to those at the others. Dr. Henstridge expects some time delay of up to 12 hours between the west coast and east coast stations; our exploratory analyses, some of which are described in Raftery *et al.* (1982), showed this not to be important at the daily level of aggregation. Professors Guttorp and Sampson detect a gradient in variance over Ireland; we agree that this is present, but it is slight and has little effect on the performance of the estimators (3.4) and (4.10).

In answer to Dr. Bhansali, (4.1) is not a special case of the standard multivariate ARMA (MARMA) model as defined, for example, by Tiao and Box (1981), because the latter does not

allow for long-memory dependence. The standard MARMA model is, however, a special case of model (A). We did some MARMA modelling at an early stage of our project (Raftery *et al.*, 1982), but this was not very satisfactory in terms of our main goal. The diagnostic checks we used are summarized in Section 4.4.

Professor Tong points out that  $E[X_{it} | X_{js}]$  could well be non-linear. We found no evidence of this in our data, but it may well be true in other situations, and model (A) could easily be modified to take account of it.

#### *Estimating d*

The discussion opposes two views about the estimation of  $d$ . Our approach, which also underlies the discussions of Dr. Carlin, Professor Künsch and Dr. McLeod, is the traditional one of exact or approximate MLE. However, Professor Dempster and Professor Smith point out that this amounts to using the fractional differencing term to shape spectra across the full frequency range, whereas a different value of  $d$  could be operating at the lowest frequencies. This leads to methods of estimating  $d$  based only on the lowest periodogram ordinates, such as those of Janacek (1982) and Geweke and Porter-Hudak (1983). Professor Smith suggests an ingenious way of making theoretical progress on the hitherto elusive properties of such methods by exploiting the analogy with the estimation of the tail of a probability distribution.

Li and McLeod (1986) and Hosking (1984a) report simulation results that MLE-type estimators perform much better than low-frequency-based estimators. Of course, this is valid only if the model fits reasonably well (and then is almost tautological), which does seem to be the case for our data. We conjecture that the ARMA terms in the model determine most of the medium and higher frequency behaviour, leaving only the low frequency behaviour to be determined by the fractional differencing term. If this is true, the problem with MLE-type estimators which concerns Professor Dempster and Professor Smith is less serious.

In clarification of remarks by Drs. Chatfield and Yar and Professor Dempster, we should say that we did not use the AR(9) residuals to estimate  $d$ , and indeed we would not want to, for

much the same reasons as Professor Dempster. Fig. 5 is used only for the exploratory purpose of revealing the presence of long-memory dependence. This may well explain the discrepancies noted by Dr. Walden, whose remarks could lead to low-frequency-based estimators of  $d$  as high as  $d = 2$ , compared with the approximate MLE  $d = 0.328$ . Values such as  $d = 2$  are incompatible with the behaviour of the sample means in Table 1.

Dr. Carlin would welcome further justification and evaluation of our approximation to the log-likelihood. Our investigations were encouraging, although of necessity somewhat limited. For example, for simulated univariate ARIMA  $(0,d,0)$  series of length 1000, we found that with  $M = 100$  the difference between our approximate log-likelihood and the exact one was generally less than the average contribution of a single observation. We intend to pursue these investigations, and we hope that others do likewise. In answer to Dr. Carlin, we used a quasi-Newton optimization method without derivatives, with starting values found as in Section 4.2.

Professor Künsch's derivation of Whittle's approximation to the log-likelihood for the model (4.1) is a real contribution, and one which we were unable to make! It is not clear that the Whittle approximation requires much less CPU time than the one we used, but we look forward to further investigation and comparison of the two approximations.

### *Asymptotics*

In Section 4.3 we said, "Neither the finite-sample nor the asymptotic distribution of the MLE for models such as (4.1) appears to be known." Dr. McLeod contests this, citing Li and McLeod (1986). However, their theorem applies only to the univariate case, and then only when the mean is known. It is thus far from yielding the distribution of the MLE for (4.1), for which there may also be problems with the nugget parameter  $\alpha$ , as pointed out by Professor Mardia. There is a further difficulty with Li and McLeod (1986). They study the univariate model

$$\phi(B) \nabla^d (X_t - \mu) = \theta(B) \varepsilon_t, \quad (D)$$

saying that  $\{X_t\}$  has mean  $\mu$ . However, it does not follow from (D) that  $\{X_t\}$  has mean  $\mu$ , since

$\nabla^d \mu = 0$ ; indeed, (D) does not specify any mean for  $\{X_i\}$ . This is why it is important to put the  $\nabla$  operator on the right-hand side of equations such as (D), as in (4.1) and (A).

We thank Professor Stein for his authoritative comments. Of course, the only sensible asymptotics in our problem refer to  $N$  large, and, as a practical matter, we accept the inapplicability of Mardia and Marshall (1984).

#### *Estimating $\mu_k$*

Dr. Beran points out that our expression (4.10) for  $\text{Var}(\bar{\mu}_k)$  does not take into account the fact that  $d$ ,  $\phi(B)$  and  $\theta(B)$  have to be estimated; this also applies to  $\alpha$ ,  $\beta$  and  $\sigma_X^2$ . However, the standard errors for these parameters appear to be small, and so it seems unlikely that taking them into account would increase  $\text{Var}(\bar{\mu}_k)$  by much. Professors Cressie and Pesarin point out that a similar comment applies to the seasonal component; we suspect that the effect of this is also small.

A more important source of variability, which we did not take into account either, is the fact that  $\mu_i$  ( $i \neq k$ ) are estimated. Because of the long-memory property, these estimates are somewhat imprecise, even with 18 years of data. Our cross-validation study was conditional on these estimates. Professor Künsch's modified estimator of  $\mu_k$  and its variance do take account of this, and are thus more realistic than our proposals. We suspect that the difference is slight in our application, but it may well be important in other contexts.

#### *Estimating wind power*

Section 5 of the paper is rather more empirical than we would prefer. In particular, extrema, while critically important to the survival of the machine, as Professor Titterington and Mr. Jamieson remark, are less important for power production, as Dr. Lippman and Professor Mollison point out. Not only will our method overestimate the machine-specific power production, if used unthinkingly, but it is probably unnecessarily pessimistic on the question of

precision. Fig. D8 helps to demonstrate Professor Lippman's point for a specific turbine, and may be contrasted with Fig. 6. The power-velocity curve relates *instantaneous* wind speed to power, and shows that the machine shuts down in high winds, for safety. Our apologies to Dr. Glaseby for his difficulties with Fig. 6; we seem to have have added a little too much 'jitter' in preparing this diagram.

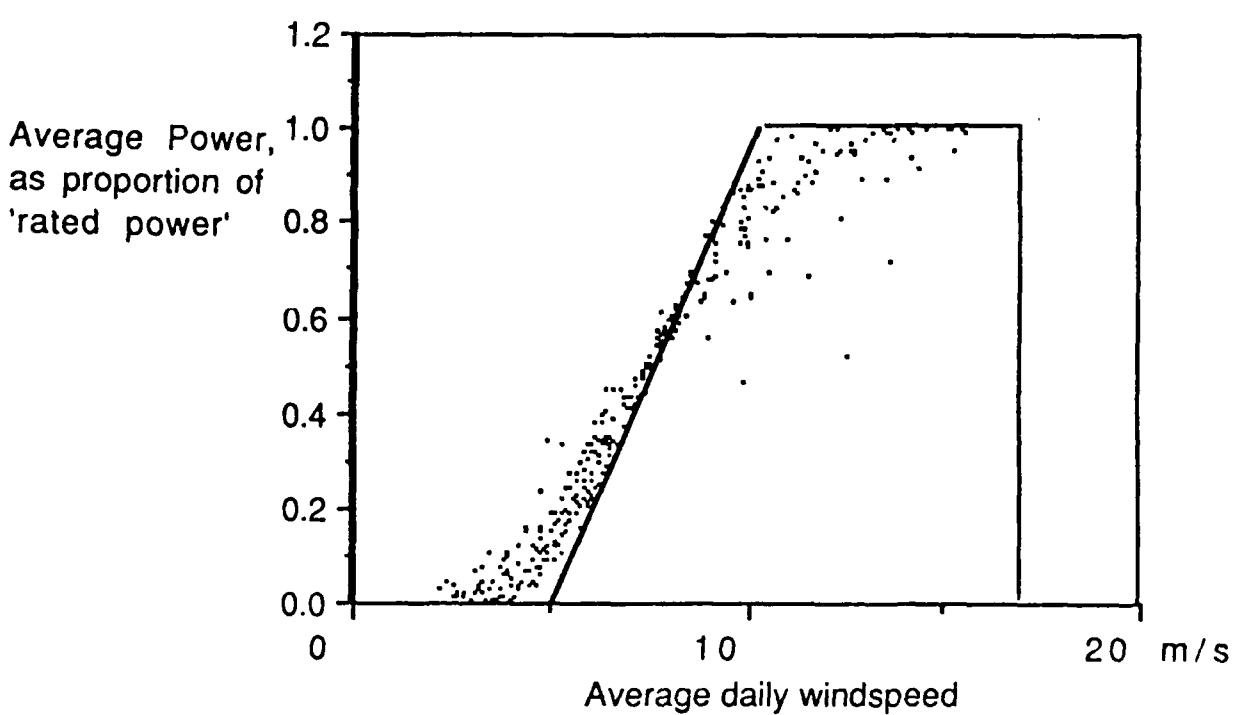


Fig. D8. Wind power generated from a given turbine, as a function of observed daily average windspeed,  $Z_{it}^2$ . The solid line is the power-velocity curve for the turbine. Note that there is no power below 5 meters per second, or above 17 meters per second. The data is for one year only at Belmullet.

It is right that Professor Mollison should remind us that there are other approaches to this problem. He mentions two: his own interesting proposal, and the meteorologically based "hindcasting" approach of Golding. We wonder how his non-parametric model could be extended to a multivariate study, with wind data at more than one site. Of course this may be less

important in studies of wave energy.

A further alternative, under development for some time at Risø, in Denmark (Peterson, Troen and Mortensen, 1988) is based on an expert evaluation of the site in question, with regard to terrain in different directions and other similar matters. It refers not only to the hourly wind data at a local synoptic station (defined by the World Meteorological Organization as a station satisfying certain exposure criteria, at which a variety of weather data are collected at least as often as every 3 hours) but also to the 'effective geostrophic wind' at the top of the boundary layer. The method yields estimates of mean wind energy, and of the distribution of wind speeds, at the chosen site, *in advance of any data at that site*. As such it provides a good example of the *a priori* information that we and Dr. Scott feel to be so important. It does not yield explicit estimates of *a priori* precision, but very recent information provided by Liam Burke suggests that a precision of  $\pm 20\%$  for mean kinetic energy has been achieved in tests at well exposed sites in Ireland. Since this can then be complemented by new data at the site, adjusted in a manner such as we have proposed, accuracy sufficient to satisfy Professor Mollison is not impossible.

#### *Miscellaneous*

Professor Künsch and Dr. Katz both cast doubt on our recommendation that windspeed data be collected at a much denser grid of locations, perhaps using simple anemometers attached to existing electricity and telephone poles. Dr. Katz's reservations are based on the debate between long memory and non-stationarity, on which we have already commented. Professor Künsch rightly points out that such information will be useful only if the records are much longer than at the site of interest; our recommendation is that they be collected permanently, if perhaps infrequently, as a supplement to the synoptic data. The question of optimally siting such new locations, or wind farms, remains, as Dr. Scott points out, an open and difficult question.

Drs. Chatfield and Yar take us to task for not smoothing the periodograms in Fig. 5. Interest there focuses on a small number of low frequency ordinates and on the narrow peak at the annual frequency, and we felt that smoothing would obscure rather than highlight these features,

which are already clear from the raw periodograms. Of course, sophisticated smoothing procedures which would not have this disadvantage are no doubt available, but using them seemed to us rather circular.

Professors Cressie and Pesarin ask whether the data are available for reanalysis. They may be obtained by sending electronic mail to Adrian Raftery at `raftery@entropy.ms.washington.edu` or `raftery%entropy.ms@beaver.cs.washington.edu`; they occupy about half a megabyte of storage.

We are grateful to Julian Besag, Liam Burke, Michael Newton, Paul Sampson and Richard Smith for helpful discussions during the preparation of this reply.

#### Additional references

Deely, J.J. and Lindley, D.V. (1981) Bayes empirical Bayes. *J. Amer. Statist. Ass.*, **76**, 833-841.

Dubey, S.D. (1967) Normal and Weibull distributions. *Naval Logistics Res. Q.*, **14**, 69-79.

Johnson, N.L. and Kotz, S. (1970) *Continuous Univariate Distributions, Distributions in Statistics*, Vols. 2, 3. Boston: Houghton Mifflin.

Madden, R.A. (1976) Estimates of the natural variability of time-averaged sea level pressure. *Monthly Weather Rev.*, **104**, 942-952.

McWilliams, B, Newmann, M.M. and Sprevak, D. (1979) The probability distribution of wind speed and direction. *Wind Engineering*, **3-4**, 269-273.

Petersen, E.L., Troen, I. and Mortensen, N.G. (1988) The European wind energy resources. To appear in *Proceedings: 1988 European Community Wind Energy Conference and Exhibition*. Denmark.

Shukla, J. and Gutzler, D. (1983) Interannual variability and predictability of 500 mb geopotential heights over the Northern Hemisphere. *Monthly Weather Rev.*, **111**, 1273-1279.

Tiao, G.C. and Box, G.E.P. (1981) Modeling multiple time series with applications. *J. Amer. Statist. Ass.*, 76, 802-816.

Trenberth, K.E. (1985) Potential predictability of geopotential heights over the Southern Hemisphere. *Monthly Weather Rev.*, 113, 54-64.

**APPENDIX**

**TEXT OF THE CONTRIBUTIONS TO THE DISCUSSION OF  
"SPACE-TIME MODELLING WITH LONG-MEMORY DEPENDENCE:  
ASSESSING IRELAND'S WIND POWER RESOURCE"**

**BY JOHN HASLETT AND ADRIAN E. RAFTERY**

## Appendix: Discussion Contributions

Professor R.L. Smith (University of Surrey), Proposer of vote of thanks

Professor Denis Mollison (Heriot-Watt University): Seconder of vote of thanks

Dr. C.A. Glasbey (Scottish Agricultural Statistics Service)

Dr. Eric Renshaw (University of Edinburgh)

Professor Knut Conradsen (Technical University of Denmark)

Dr. I.T. Jolliffe (University of Kent)

Dr. C. Chatfield and Dr. M. Yar (University of Bath)

Dr. J.T. Kent (University of Leeds)

Mr. P.B. Brontë-Hearne

Dr. R.J. Bhansali (University of Liverpool)

Professor Toby Lewis (University of East Anglia)

Dr. Jan Beran (Texas A & M University)

Dr. J.B. Carlin (La Trobe University)

Professors N.A.C. Cressie (Iowa State University) and F. Pesarin (Università degli studi di Padova)

Professor A.P. Dempster (Harvard University)

Professors Peter Guttorp and Paul D. Sampson (University of Washington)

Dr. John Henstridge

Dr. D.A. Jones (Institute of Hydrology)

Dr. Richard W. Katz (National Center for Atmospheric Research)

Professor Hans R. Künsch (ETH Zurich)

Dr. W.K. Li (University of Hong Kong)

Dr. Alan Lippman (Brown University)

Professor K.V. Mardia (University of Leeds)

Dr. A.I. McLeod (University of Western Ontario)

Dr. Yoshihiko Ogata (Institute of Statistical Mathematics)

Dr. W.D. Ray (Birbeck College)

Dr. E.M. Scott (University of Glasgow)

Professor Michael Stein (University of Chicago)

Professor Paul Switzer (Stanford University)

Dr. Winson Taam and Professor Brian S. Yandell (University of Wisconsin-Madison)

Professor D.M. Titterington (University of Glasgow) and Mr. P. Jamieson (James Howden)

Professor H. Tong (University of Kent at Canterbury)

Dr. Andrew Walden (B.P. Exploration)

## RESEARCH SECTION PAPER BY HASLETT AND RAPTERY

Vote of Thanks proposed by  
Professor R.L. Smith  
University of Surrey

This paper is an excellent example of the development of statistical methodology to solve a substantial applied problem.

The problem is typical of those which arise in what may loosely be termed the environmental sciences - by these I include such fields as hydrology, meteorology, air pollution and numerous problems with a biological flavour. As such, the methods used will be of interest to workers in all these fields.

The authors' approach incorporates many techniques. After initial exploratory analysis they propose a "kriging" estimator for interpolation at a new site, exploiting spatial correlations. Further analysis leads them to identify a model incorporating long-range and short-range temporal correlations. The method of fitting, based on an approximate likelihood function, makes an original contribution to the computational aspect of time series models, and finally the model is applied, not without further difficulties, to the prediction of wind power.

In seeking some aspect on which to comment in more detail, my attention naturally fell on the long-memory aspects, which of all the authors' techniques are the ones least well understood at the moment. I therefore went back to the last time a paper before this Society was substantially concerned with this theme, Lawrence and Kottekoda (1977), and found the following quotation:

"Long-term dependence has in the past been analysed using the rescaled adjusted range...; the method has been propounded by Mandelbrot and Wallis... and so far it has no competitors."

The rescaled adjusted range has not been nearly so prominent in the recent literature of this subject. Why did the method become fashionable,

and why did it become unfashionable again?

Part of the reason, no doubt, lies in the introduction of the fractional differencing concept. Although there have been many theoretical papers on this subject, there are few containing really substantial applications, and tonight's paper is to be welcomed if only for that reason.

Nevertheless, this approach is very much model-dependent. The analyst who is uncertain whether to use a long-memory model at all may well prefer a nonparametric, robust approach to the estimation of  $d$ . One such has been proposed by Geweke and Porter-Hudak (1983). Assuming a spectral density of the form

$$f(\lambda) = O(\lambda^{-2d}), \quad \lambda \rightarrow 0,$$

their method is based on the approximate linearity of  $\log I_N(\lambda)$ , the log of the periodogram based on  $N$  observations, in  $\log \lambda$ . Roughly, they fit a least-squares linear regression to  $\log I_N(\lambda_j, N)$  against  $\log \lambda_j, N$  for  $j=1, 2, \dots, n$  ( $\ll N$ ), where  $\lambda_j, N = 2\pi j/N$  is the  $j$ 'th Fourier frequency, and estimate  $-2d$  as the slope of that regression.

This approach has some analogies with estimating the tail of a probability distribution. For example, under an assumption of the form

$$f(\lambda) = a\lambda^{-2d} (1 + b\lambda^c + o(\lambda^c)), \quad c > 0,$$

one can show that the optimal  $n$  is of order  $N^{2c/(2c+1)}$ , with corresponding mean squared error of order  $N^{-2c/(2c+1)}$ . The calculation mimics Hall (1981) in the tail estimation context; Hall and Welsh (1984, 1985) have considered some other aspects of this.

There is a technical difficulty with this calculation; namely, that the standard sampling properties of the periodogram (approximately independent and exponentially distributed ordinates at the Fourier frequencies) fail in the extreme lower tail under a long-memory model. This is also a technical gap in the paper of Geweke and Porter-Hudak, and may well have something to do with the levelling-off of the periodogram in the extreme lower tails of the authors' Figure 5.

Returning to the methodological aspects of the paper, in view of the wide range of potential applications I think it is worth examining some of the assumptions from a broader viewpoint than just whether they were justified for this particular data set. I had some doubts about both equation (3.3), where there is no allowance for any kind of directional dependence, and the constancy of ARMA coefficients across all sites in Section 4.1. Do the authors have any comments on whether such assumptions are likely to prove restrictive in trying to apply the model in other contexts? What alternatives are available?

Overall, this paper must be praised as a major piece of applied work, for the development of new methodology, for its contribution to the computational aspect of long-memory model fitting, and not least for the theoretical developments it will stimulate. It is an ideal contribution to the proceedings of this Society.

I do not know whether the authors feel that Irish statistics have been neglected by this Society in the past, but Dr. Haslett did take the trouble to remind us, in his presentation tonight, where Ireland is. I am sure that we would all hope that the Irish winds will blow some more papers over to us, and that that process, at least, is one that will not require from us a long memory. I have great pleasure in proposing a vote of thanks.

#### REFERENCES NOT IN THE TEXT

Hall, P. (1982), On some simple estimates of an exponent of regular variation. *J.R. Statist. Soc. B* 44, 37-42.

Hall, P. and Welsh, A.H. (1984), Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* 12, 1079-1084.

Hall, P. and Welsh, A.H. (1985), Adaptive estimates of parameters of regular variation. *Ann. Statist.* 13, 331-341.

Lawrance, A.J. and Kottekoda, N.T. (1977), Stochastic modeling of river-flow time series (with discussion). *J.R. Statist. Soc. A* 140, 1-47.

Seconding of vote of thanks at RSS Meeting, 25 May 1988 (Haslett & Raftery)

---

Professor Denis Mollison (Heriot-Watt University, Edinburgh): Where Richard Smith has discussed the theoretical content of tonight's paper, I shall concentrate on the applied side. The problem addressed by the authors is indeed of practical importance, and their conclusion is somewhat depressing: even with nearly a year's data from a new site ( $n = 320$ ), confidence intervals for the mean resource have a  $\pm 30\%$  spread (Table 2), where we might have assumed an accuracy 4 to 5 times as great before they pointed out the importance of long-term memory dependence (see Table 1 *et seq*). Errors of this magnitude ( $\pm 30\%$ ) would affect the unit cost of wind power by about  $\pm 20\%$  (Anon 1987), which could be crucial for a resource which is on the verge of economic viability.

The authors have mentioned possible improvements in accuracy based on the use of the same data set, such as the use of Bayesian priors. An alternative, exploiting our understanding of atmospheric dynamics, would be to use a hindcasting model such as that of the UK Met Office (Golding 1980), which has produced estimates for an approximately 50 km grid covering NW Europe including Ireland since about 1978. Short period measurements for a specific site could be used to calibrate estimates from such a model, which might first be modified to take account of local topography.

In the other direction, an alarming possibility is that the wind climate may be appreciably non-stationary on the time scale considered (say 10 to 50 years). Carter and Draper (1988) have recently pointed out strong evidence for a significant increase in wave power for sites south and west of Ireland, possibly as large as a doubling of the mean resource over the period 1960-90. Admittedly they did not detect a significant change in wind climate at the sites they considered, but since waves are generated by winds (mainly non-local, see e.g. Mollison 1986) their work certainly implies that similarly significant changes could also occur in the wind power resource.

A small point, but of some importance, is that the seasonal variation has been assumed to be the same at all sites. It would be interesting to know if the authors investigated this, and whether their conclusions might be sensitive to this assumption.

The authors' main model, with long-term memory, is in the end only used for confidence intervals. The estimator itself turns out to be in reasonable agreement with their earlier estimator, which they therefore fall back on. The latter is essentially an average of the short-term data weighted according to their simpler 'inverse-covariance' model (eq. 3.4).

This encourages me to describe a model of my own (Mollison 1980) for a similar problem, the augmentation of short-term data on wave power by longer term wind information. The approach was rather different, but there are sufficient similarities that each may illuminate the other. My approach was initially based on a model for a wave power measure  $P_i$ , the average power observed in month  $i$ , in terms of a predictor based on the average value of the fifth power of wind speed,  $W_i$ .

~~Power~~  
$$\ln(P_i) = k + \ln(W_i) + c_i$$

Like the authors' equation (3.4) this is a linear relation between transformed values of short-term and long-term variables.

This parametric model yields estimates  $P_i$  for the longer period, and in particular an estimate and confidence interval for the mean wave power resource. For instance, with wave data for two years ( $n = 24$ ) and wind data for 13 years ( $N = 156$ ), the confidence interval was estimated at  $\pm 13\%$ . However, results were sensitive to the details of the model; the estimates  $P_i$  ranged up to more than twice the highest observed value, and thus the estimate of the mean resource was sensitive to the power of windspeed used in defining  $W_i$ .

A nonparametric alternative is to assume only that  $P$  depends monotonely on  $W$ . If this is the case, we can estimate the distribution function of  $P$  using all the values of  $W$  to determine the vertical scale; that is, we plot  $P_i$  against the position of  $i$  among the order statistics of  $\{W_i\}$  (see Figure). A non-decreasing estimate of the distribution function can be ensured by a monotone least squares regression (dotted line in Figure).

This method has a number of advantages, apart from its minimum of assumptions. There is no need to estimate the relational parameter  $k$ , which is the main contributor to the uncertainty in our estimate of the mean resource; so it is not surprising that there is little if any loss of accuracy in the estimate of the mean resource. Indeed, simulations for my particular data set, admittedly with a slightly different treatment of the highest end of the power range, actually gave a narrower confidence interval,  $\pm 10\%$ , than for the parametric model.

Perhaps the greatest advantage of the nonparametric method, however, is that it can be interpreted as giving weights to the short-term data: namely, data month  $i$  is given weight proportional to the number of months in the ordered sequence  $(W_{i,j})$  for which it is the closest data month. (A slight refinement is to share out weights equally where data months are in the wrong order, that is among months for which the monotone least squares regression mentioned above takes the same value. Simulations suggest that this also slightly increases the accuracy of the estimate of the mean resource.)

The complete set of short-term data can then be used, with these weights, as a representative resource sample; for instance, in the wind and wave power contexts such a set can be used to optimise device design (see, e.g., Mollison 1980). There should be no difficulty in extending this representation to the authors' case of a number of synoptic stations; their equation (3.4) essentially gives weights to the various synoptic stations, and thus could be used to combine sets of weights derived as above for the individual stations.

~~The nonparametric method may fail to represent extreme conditions, especially in a sample where there are few observations in what, on the evidence of the background data  $W_t$ , were the most extreme months. I would argue that this is actually an advantage, in that it makes it clear that we do lack this information; it is precisely in these circumstances that we would be unwise to rely on the parametric model. In particular, it indicates that where extremes are of interest, as in design survival tests, further data or different estimation techniques are required. On the other hand, knowledge of extremes is unnecessary for power output estimates, since almost by definition they will be beyond the output limit of economic devices.~~

There remains the problem of long-term memory. Even taking monthly averages, the sequence  $(W_t)$  showed a (seasonally detrended) serial correlation of 0.2. In the light of the authors' analysis, it would clearly be desirable to reassess my estimates of confidence intervals.

The methodology of tonight's paper has of course much wider generality than applications to renewable energy; but it is applications such as this which motivate developments in the methodology, and John Haslett and Adrian Raftery's exposition balances the interest of the two in a way that is most welcome. It deserves to remain in our long-term memory, and I have much pleasure in seconding the vote of thanks.

#### References

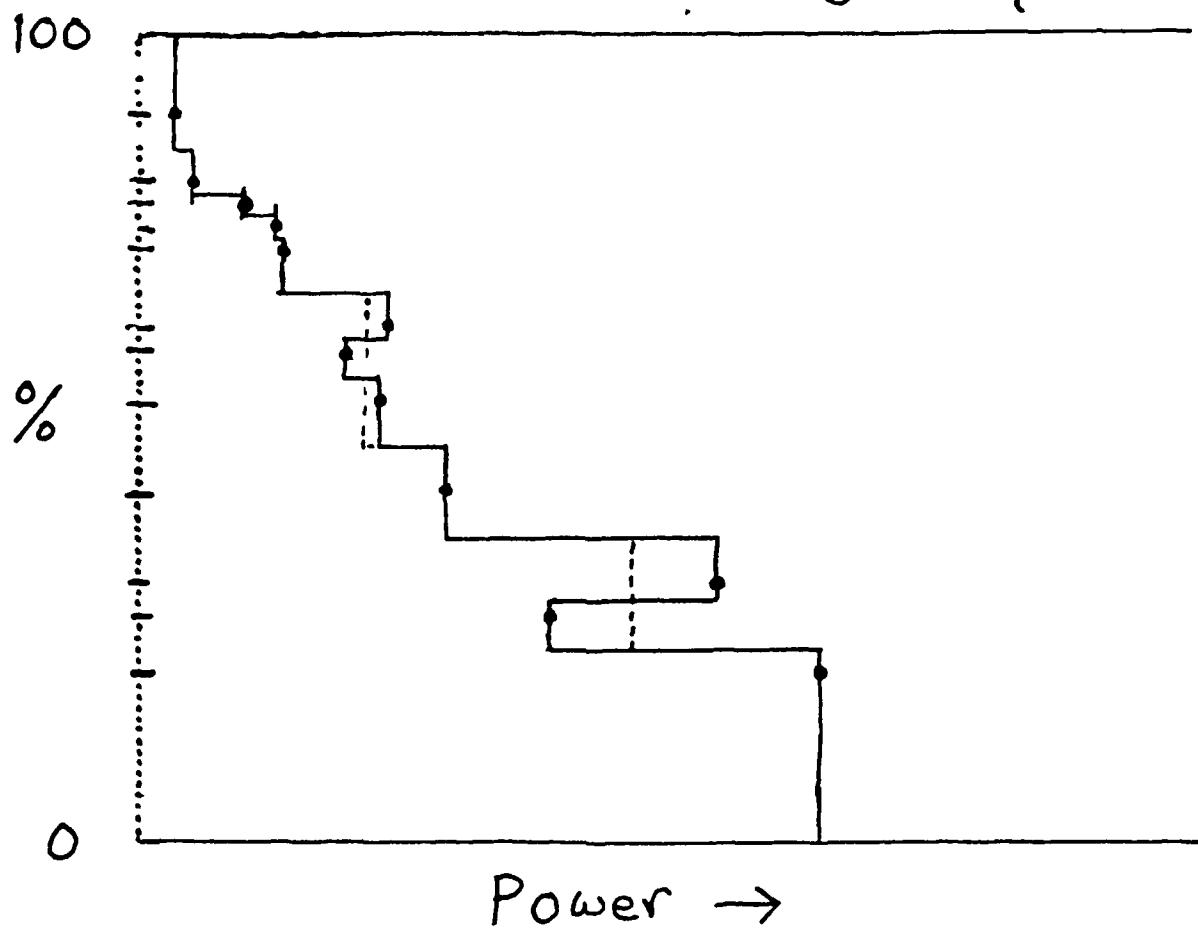
Anon (1987) *Wind Power for the UK*. London: British Wind Energy Association.

Carter, DJT & Draper, L (1988) Has the north-east Atlantic become rougher? *Nature* 332, 494.

Golding, B (1980) Computer calculations of waves from wind fields. In *Power from Sea Waves* (B Count, ed), pp 115-134. London: Academic Press.

Mollison, Denis (1980)  
Mollison, Denis (1986) } both in refs to main paper

augmented data using order  
within longer sequence



Note interpretation as giving  
weights to short term data

→ Can combine weights from  
a no. of seqs. ( $w_i^{(t)}$ ) (eq. 3.4)

Comments on paper by Haslett and Raftery

Dr C.A. Glasbey

Scottish Agricultural Statistics Service,  
JCMB, The King's Buildings, Edinburgh EH9 3JZ

I enjoyed this paper, which is an attractive blend of theory and practice, and a good example of the usefulness of statisticians.

A principal components analysis of the spatial covariance matrix gives an alternative perspective on its structure. Based on  $R$ , 80% of the spatial variability is accounted for by a daily average, and half of what remains by a linear gradient across Ireland. By way of comparison, I am involved with the Scottish Centre of Agricultural Engineering in studying local variability in solar radiation in the Pentland Hills, to the South of Edinburgh. We have also found a square-root transformation to be appropriate for stabilising variances. In our case,  $3/4$  of the spatial variability about a daily mean is explained by a linear gradient. Most of this variability is concentrated in a few days when either a north/south, or an across-the-ridge, effect occurs.

Meteorologists have a rule-of-thumb that about 30 years of weather data is optimal to represent current climatic variability, because longer periods are affected by drifts in climate. Arising out of this, how does long-term memory relate to climatic drift? And, would the authors have used 100 years of data if they had had them available?

Have the authors considered the possibilities which exist, for larger values of  $n$ , of increasing the robustness of inference. For example, elements in row  $k$  of  $R$  could be estimated, to guard against the 1 in 12 chance of being at another "Rosslare"! Equation (5.3) looks highly sensitive to the normality assumption. An estimator constructed by resampling the data may perform better.

Two points of detail: I could not understand how it is possible that some of the data points in Fig. 6 correspond to  $\sqrt{3} < z^6$ , and the results in Table 2 look unexpectedly good. If log-normal approximations are used and small correlations ignored, then the squared distance between the vectors of point and "true" estimates is about 6. This lies in the lower 10% tail of a  $\chi^2_{12}$  distribution!



ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

1. DEPARTMENT OF STATISTICS, J.C.M.B.  
KING'S BUILDINGS, MAYFIELD ROAD,  
EDINBURGH EH9 3JZ.

2. Name (incl. title)

DR. ERIC RENSHAW

3. Affiliation (as you wish it to appear on your printed contribution)

UNIVERSITY OF EDINBURGH

Text of Contribution (Double spaced)

(PLEASE SEE ATTACHED SHEET)

The authors are to be congratulated for presenting such a stimulating array of theoretical and practical aspects of space-time modelling. Long-range memory processes are of particular importance, and insight into their behaviour can be obtained by considering spatial persistence through the interaction model (Barlett 1971, 1975)

$$X_{ij}(t+dt) = X_{ij}(t)(1+dt)^{\gamma} + \sum_{r,s} a_{rs} (X_{ij}(t) - X_{i+r, j+s}) dt + dZ_{ij}(t) + \sigma(dt)$$

where  $\{X_{ij}(t)\}$  is a lattice-process on  $i, j = \dots, -1, 0, 1, \dots$  and  $\{dZ_{ij}(t)\}$  denotes white-noise. The associated spectrum (Renshaw, 1984) for equal  $\{X_{ij}(0)\}$  is

$$f(\cdot_1, \cdot_2; t) = (\cdot^2/\cdot) [\exp(\cdot t) - 1]$$

$$\text{where } \cdot(\cdot_1, \cdot_2) = 2 \{ \sum_{r,s} a_{rs} [1 - \cos(r_1 t + s_2 t)] \}$$

and this is especially useful for seeing how the form of the interaction weights  $\{a_{rs}\}$  determines overall spatial structure.

For example, the one-dimensional case with nearest-neighbour weights  $a_{ij} = a_{j-i}$ ,  $a_j \neq 0$  (otherwise) leads to  $\cdot = 2 + 8 \cdot \sin^2(\cdot)$ . If  $\cdot = \pi/4$  then  $\cdot = 0$  for  $t = 0$  where  $0 = \cos^{-1}[1/(2\cdot)]$ , whence  $f(\cdot; t) \sim \cdot^{-2/3}$  as  $t \rightarrow 0$ . Thus  $\cdot$  defines an "inter-scale of pattern" in the sense that if  $\cdot \rightarrow 0$  then  $f(\cdot; t)$  does not approach a stationary limit as  $t \rightarrow \infty$ . So changing  $\cdot$  enables us to alter the range of the scales of pattern present in the stationary part of the process. If  $\cdot \neq 0$  then for small  $\cdot$  we have the inverse-square law

$$f(\cdot; t) \sim -(\cdot^2/2\cdot) \cdot^{-2} \quad (\cdot \neq 0)$$

whilst the Cauchy-type weights  $a_r = k/[(\cdot r)^{(\cdot r+1)}]$  ( $r \neq 0$ ) yield pure "1/ $\cdot$ -noise", viz.

$$f(\cdot; t) \sim (\cdot^2/2k^2) \cdot^{-1}$$

Interest in spatial persistence requires us to extend this by constructing a process which possesses a genuine power-law spectrum  $f(\omega; t) \sim \text{const.} \omega^{-d}$  for non-integer  $d > 0$ . This may be achieved by using a similar fractional differencing approach to the authors. For the ARIMA  $(0, d, 0)$  process  $x_t = (1 - B)^{-d} x_t$  yields negative binomial weights which suggests putting  $a_r = c \left[ \frac{r^{r+d-1}}{r!} \right]$  ( $r \geq 0$ ). These give rise to

$$\psi(\omega) = 2\pi - 4c[2\sin(\pi\omega)]^d \cos\{\pi(\pi-\omega)d\} \quad ,$$

and so

$$f(\omega; t) \sim [\pi^2/4c \cos(\pi\omega d)]^{\omega-d} \quad (\text{if } \lambda=0)$$

as required.

#### References

Bartlett, M.S. (1971) Physical nearest-neighbour models and non-linear time-series. *J. Roy. Statist. Soc., B*, **8**, 222-232.

Bartlett, M.S. (1975) *The statistical analysis of spatial pattern*. London: Chapman and Hall.

Renshaw, E. (1984) Competition experiments for light in a plant monoculture: an analysis based on two-dimensional spectra. *Biometrika*, **71**, 717-728.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

---

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

1. IMSOR, Building 321,  
Technical University of Denmark  
DK 2800 Lyngby, Denmark

2. Name (incl. title) Professor Knut Conradsen,

3. Affiliation (as you wish it to appear on your printed contribution)

The Technical University of Denmark

---

Text of Contribution (Double spaced)

I would like to congratulate the authors on a stimulating paper that in an impressive way applies recent developments in time series and spatial statistics in the analysis of large data sets. The paper clearly demonstrates the importance of involving statisticians in work that otherwise often are done exclusively by physicists and engineers.

My comments relates to the problems around the spatial interpolation. In geostatistics one applies different types of Minimum Mean Squared Error estimates based on different models for the spatial autocovariance. It is common folklore that the results of such an interpolation (a so-called kriging) are fairly insensitive to some misspecifications of the spatial covariance structure, cf. the remarks following (3.4).

In figures D1 and D2 is shown the kriging variance and the kriging weights in a simple kriging problem with 3 observations. The semivariogram is spherical with nugget effect  $c_0$  and sill  $c_0 + c_1$ . We see that the kriging weights are fairly sensitive to changes in the relative nugget effect  $c_0/(c_0 + c_1)$ . Our experience working with geochemical samples (stream sediments) has been that this may have very serious effects whenever the data structure deviates from the model. In this sense, I do not think that one should consider kriging to be a fairly robust technique.

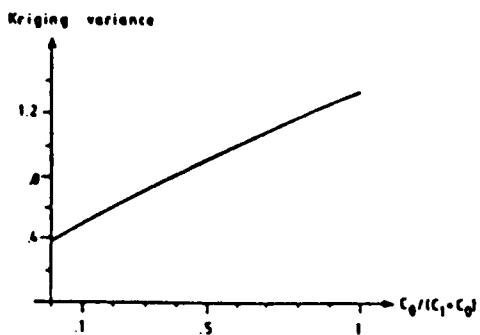


Fig. D1.

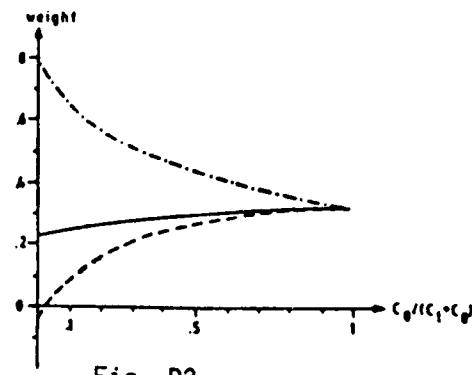


Fig. D2.

My second remark is related to the first, namely the question of a proper modelling of the spatial autocovariance. The authors have chosen the exponential given in (3.3). In the interpolations the behaviour of the autocorrelation close to 0 is very important. In the region say between 0 and 50 kms I do not, however, think that the fit offered by the authors is very adequate. A closer scrutiny of figure 3 shows that the correlations between 60 and 100 kms vary around 0.87, with no systematic decrease in that region. From two danish meteorological stations with a distance of only 6 kms a correlation of 0.87 was found (based on 7500 observations). If we add this observation and reestimate the correlation structure, the outcome could be as in figure D3.

In actual interpolations this could be of importance. The model checking in the paper is based on a cross validation technique, and therefore only correlations between sites with larger differences are used. It will, of course, be trivial to modify the correlation structure, and my remarks shall only serve the purpose of pointing out some possible pitfalls in modelling spatial data.

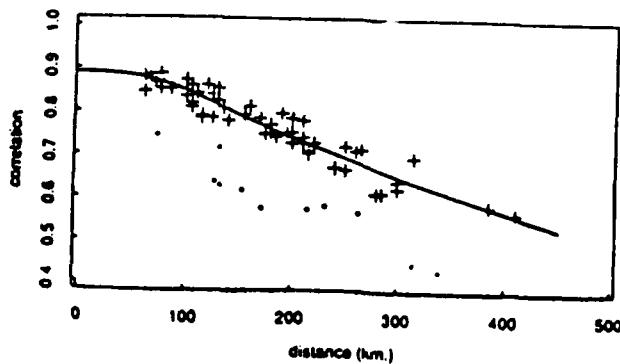


Fig. D3.

Contribution to the Discussion of

"Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource" by J. Haslett and A. E. Raftery

---

Dr. I.T. Jolliffe (University of Kent)

I would like to thank the authors for a simulating paper, which uses, in an interesting way, some relatively recent ideas from Time Series Analysis and Spatial Modelling on a real data problem. I have three comments, two of which relate to the somewhat strange behaviour of the data from the station at Rosslare. Without knowing anything about the siting of the station, it would seem to me more likely that the difference between it and the other stations is due to local topography rather than to a regional effect. The main part of the discrepancy noted in the paper between Rosslare and the other stations is in the inter-station correlations (figure 3), but it may be that it is the different auto-correlation structure at Rosslare (figure 4) which is the more fundamental difference. Consider the following (oversimplified) model involving two stations only.

Let  $\epsilon_{1t}$ ,  $\epsilon_{2t}$  be the noise terms for the two stations, each with variance  $\sigma_\epsilon^2$  and with

$$\text{corr}[\epsilon_{1t}, \epsilon_{2t}] = \rho_\epsilon.$$

Suppose that the velocity measures  $x_{1t}$ ,  $x_{2t}$  follow AR(1) models

$$x_{1t} = \phi_1 x_{1t-1} + \epsilon_{1t}$$

$$x_{2t} = \phi_2 x_{2t-1} + \epsilon_{2t}$$

Then  $\text{var}[x_{it}] = \sigma_\epsilon^2 (1 - \phi_i^2)^{-1}$ ,  $i = 1, 2$  and

$$\text{cov}[X_{1t}, X_{2t}] = \sigma_e^2 \rho_e (1 - \phi_1 \phi_2)^{-1}.$$

so the correlation between  $X_{1t}$  and  $X_{2t}$  is given by

$$\rho_X^2 = \rho_e^2 \frac{[(1 - \phi_1^2)(1 - \phi_2^2)]}{(1 - \phi_1 \phi_2)^2} = K \rho_e^2, \text{ say.}$$

Now  $K \leq 1$ , and the amount by which  $\rho_X^2$  is shrunk relative to  $\rho_e^2$  depends on the difference between the denominator and numerator of  $K$ , namely  $(\phi_1 - \phi_2)^2$ . There is no shrinkage when  $\phi_1 = \phi_2$ , but as  $\phi_1, \phi_2$  diverge, so shrinkage increases. Thus, the smaller cross-correlation for Rosslare may be an indirect effect of smaller auto-correlation. I would welcome the authors comments on this.

The second question regarding Rosslare is to ask whether the cross-validation exercise has been extended to predict the values for Rosslare. If the results are reasonable for this atypical site, it would increase confidence that worthwhile predictions can be made at new sites.

My final point is a brief question concerning the model (4.1). The authors allow any past dependence of one  $X_{it}$  series on another to be explained entirely in terms of correlation between noise terms. To what extent is this less flexible than allowing direct dependence of  $X_{it}$  on  $X_{jt-1}$ , say, for  $i = j$ ?

Dr C. Chatfield and Dr M. Yar (University of Bath)

The authors are to be congratulated for tackling such an important practical problem and presenting a paper combining so many interesting theoretical and practical topics. Given the mammoth nature of the project, the authors have done well to restrict the length of the paper to 19 sides but they have inevitably had to leave out some details, and our comments are mostly in the nature of questions to clarify a few obscurities.

First we think a footnote defining "synoptic" would avoid everyone having to look it up in the dictionary (and our dictionary didn't help much!). Secondly, a brief description of "kriging" would prevent many readers from feeling ignorant. As we understand it, kriging is a two-dimensional interpolation and smoothing method, used in the mining industry, which is related to spline smoothing (e.g. see Wegman and Wright, 1983). Our third minor query is to ask why Figure 5 presents periodograms rather than smoothed spectra which might be easier to interpret. A common vertical scale might also assist comparisons.

Our main query concerns equation (4.1) which assumes that the *same* univariate model is appropriate at each site, with the *same*  $\phi$  and  $\theta$ . We would like further justification of this assumption. We are also puzzled because in Section 4.2 the model appears to be fitted, not to the  $X$ 's (as implied by equation (4.1)), but to the fractionally differenced filtered  $Y$ 's. As we understand it the same AR filter of order 9 and the same  $d$ -value is used for each series. How was the AR filter selected and what form does it take? This is one of the first reported cases of fractional differencing that we have seen, and we would also like to see further justification of this aspect. It is not obvious to us why the more usual differencing with an integer  $d$ -value is not used. We suspect that fractional differencing arises from the shape of the (filtered?) spectrum near zero frequency, and that  $d$  is constrained to lie within the interval  $[0, \frac{1}{2}]$  in order to get a finite variance.

- 2 -

Looking at Figure 4, our first reaction was that there are substantial differences in the behaviour of the ac.f. at different sites and that it is hard to see "striking similarities between its pattern and extent at the different stations" as suggested by the authors. At Rosslare, for example, the autocorrelations are "small" at lags 5 or more and we see no need for any kind of differencing. However, at Clones, the ac.f. does not damp down to zero even at lag 100 and our first reaction is to take first differences, rather than fractional differences. No doubt this is partly due to our lack of familiarity with fractional differences, but it is certainly true that we find them difficult to interpret. A model for simple differences is easier to fit and to understand. If the same seasonal filter was used on each of the raw data series, we also wonder if some of the long-term persistence could be induced by the imperfect nature of the seasonal filter. Returning now to the periodograms in Figure 5. we find it hard to say whether they have similar properties or not (see our earlier comment on presentation). Of course as the short-memory variation has been removed from each series, the periodograms are bound to look fairly similar in that variation is concentrated at low frequencies.

The final step in Section 4.2 says that a common ARMA model is identified for all the  $\{\nabla^d Y_{it}\}$ , but gives no indication how this is done. Was an AR(2) model identified for every single site, and, if not, how were the disparities between the selected models resolved?

#### Reference

Wegman, E.J. and Wright I.W. (1983). Splines in Statistics. J. Amer. Statist. Ass., 78, 351-365.

Contribution to paper by Haslett and Raftery

read to the RSS, May 25th, 1988.

Dr. J.T. Kent (University of Leeds): I would like to congratulate the authors on a masterly application of ideas from spatial analysis, time series and long-range correlation to an important practical problem. My comments are directed to the initial data processing, which appears to consist of 3 steps.

- (a) Start with the hourly average wind speed,  $U(t)$  say.
- (b) Calculate daily averages,  $\bar{U}(t)$ , say.
- (c) Make a power transformation  $\bar{U}(t)^\alpha$ , with  $\alpha = \frac{1}{2}$ , to produce an approximate Gaussian time series.

Here are my comments.

1. Does the choice of power  $\alpha = \frac{1}{2}$  depend on the scale of temporal aggregation; that is would  $\alpha = \frac{1}{2}$  still be appropriate if weekly or monthly averages were used instead of daily averages? Related considerations arise in mining where lognormal spatial processes (corresponding to  $\alpha = 0$  above) are observed. It is found that, to a good approximation, lognormality often persists over several scales of spatial aggregation; see e.g. Dowd (1982).
2. If we also take account of the average hourly wind direction then  $U(t)$  can be regarded as the radial component of a two-dimensional wind velocity vector  $V(t) = (V_1(t), V_2(t))$ . The simplest model for the marginal distribution of  $V(t)$  is bivariate normal with mean 0 and isotropic covariance matrix, so that  $U^2(t)$  is proportional to a  $\chi^2_2$  variate. The

Wilson-Hilferty transformation of  $U(t)$  to achieve approximate normality corresponds to  $\alpha = \frac{2}{3}$ . Further if the mean of  $V(t)$  is non-zero we would expect a choice of  $\alpha$  nearer to 1. Thus the fact that the preferred choice  $\alpha = \frac{1}{2}$  is smaller than  $\alpha = \frac{2}{3}$  suggests, perhaps not surprisingly, that the distribution of  $V(t)$  is more heavily-tailed than the normal distribution.

3. Steps (b) and (c) can be carried out in either order; that is we might transform before taking averages. Indeed we might have defined the initial data  $U(t)$  to be the hourly average of wind speed to some power rather than of wind speed itself, especially as it is the cubed wind speed which is proportional to energy. Can the authors give some insight into their preferred ordering of steps?

#### References

Dowd, P.A. (1982) Lognormal kriging - the general case. J. Mathematical Geology 14, 475-499.

MR P.B. BRONTE-HEARNE: I can echo what the previous speakers have said. It was a very interesting paper. I was particularly interested in the kriging estimator. What I intended to say has already been said by other discussants.

like to  
The only thing to which I would draw attention is the power law equation. In finding a site for a wind turbine generator there has to be a certain relationship between the type of device that will be fitted and the power law equation. When there is a mean wind speed at a certain height there has to be a relationship between the height of the wind turbine generator and the mean wind speed at that particular height.

If we have a simple formula, where  $V_Z$  is the mean wind speed at height  $Z$  and  $V_H$  is the mean wind speed at height  $H$  which, for normal purposes, is approximately 10 m. the wind speed varies considerably according to the type of ground chosen. Thus  $n$  is a variable, varying with the nature of the terrain, and that will also have an effect on the suitability of the wind turbine generator.  $n$  can range from 0.1 (sand and ice) to 0.25-0.4 (for urban terrains) but is generally taken to be  $1/7$ .

display formula here

$V_Z = V_H \left(\frac{Z}{H}\right)^n$

\* see attached notes

Contribution to the discussion of the paper

by Haslett and Raftery

Dr. R. J. Bhansali (University of Liverpool):

I would also like to congratulate the authors on an interesting and substantial empirical study. I have two brief questions: First, what checks did the authors make, apart from plotting the log-periodogram against the logarithm of frequency, before deciding that they are indeed dealing with a long-memory model ? Parzen(1983) has proposed an index for diagnostic checking of long-memory models. Are the authors aware of Parzen's work and have they experience of using this index ?

Secondly, the spatial time model (4.1) considered by the authors may be viewed as a special case of a multivariate ARMA model. Have the authors tried to subject their data to the standard multivariate ARMA model fitting exercise and, if so, what sort of results did they find ? Were they totally discouraging ?

Reference

Parzen, E. (1983). Time series model identification by estimating information, memory and quantiles. Technical Report, Department of Statistics, Texas A & M University, USA

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

---

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)	1. <u>39 Walcot Road</u> <u>DISS, Norfolk IP22 3DH</u>
2. Name (incl. title)	<u>Professor Toby LEWIS</u>
3. Affiliation (as you wish it to appear on your printed contribution)	<u>University of East Anglia</u>

---

Text of Contribution (Double spaced)

Professor Toby Lewis (University of East Anglia): May I add my congratulations to the authors on a highly effective use of statistical methodology in the service of an important social need. I have a couple of tangential comments on aspects of the model.

First, regarding wind direction, there was the surprising observation in Section 6 that, when wind speed at each station was decomposed into components parallel and perpendicular to the prevailing wind direction, the relation between inter-station correlation and distance  $d_{ij}$  disappeared. I do not know whether the correlations were calculated from signed components  $v_i \cos(\sin)\theta_i$ , absolute components  $|v_i \cos(\sin)\theta_i|$ , or square roots; in any case the non-dependence on  $d_{ij}$  seems counter-intuitive. Would the authors tell us a bit more? *green V.L.L.*

Secondly, a comment on Fig.3 (which I offer in the spirit of "lateral thinking"). The model (3.3) for  $r_{ij}$  in terms of  $d_{ij}$  fits well, but there is an outlier, Rosslare, already discussed by Dr Jolliffe and other speakers: the correlations involving Rosslare are too low. However, one might equally say that the distances to Rosslare are too short! Take for instance point P,

i.e. (Dublin, Rosslare), in Fig.A below. The distance from Dublin to Rosslare is only OP, but one would like it to be OQ, right up to the fitted curve. Then why not move Rosslare? If we draw circles on the map with centres such as Dublin and radii such as OQ, the desired new location for Rosslare emerges. In the spirit of Anglo-Irish entente (and may I echo earlier speakers and say what a pleasure it is to have our friends from Dublin addressing the Society this evening), the new location proves to be in England - just. It is at Hartland Point on the north Devon coast (Fig.B below). Replotting the eleven Rosslare correlation points in Fig.3 with distances adjusted to Hartland Point we get the points  $\odot$  in Fig.A, now lying comfortably on or near the fitted relationship. Incidentally, the points R and S for Belmullet and Malin Head, lying a little off the fitted curve, could be brought nicely on to it if we shifted Rosslare, not to Hartland Point, but to the location marked  $\ast$  on Fig.B. This is the Devon village of Sheepwash. But I feel that I should stay with Hartland Point, as more fitting to the gravitas of Dr Haslett and Dr Raftery's admirable paper.

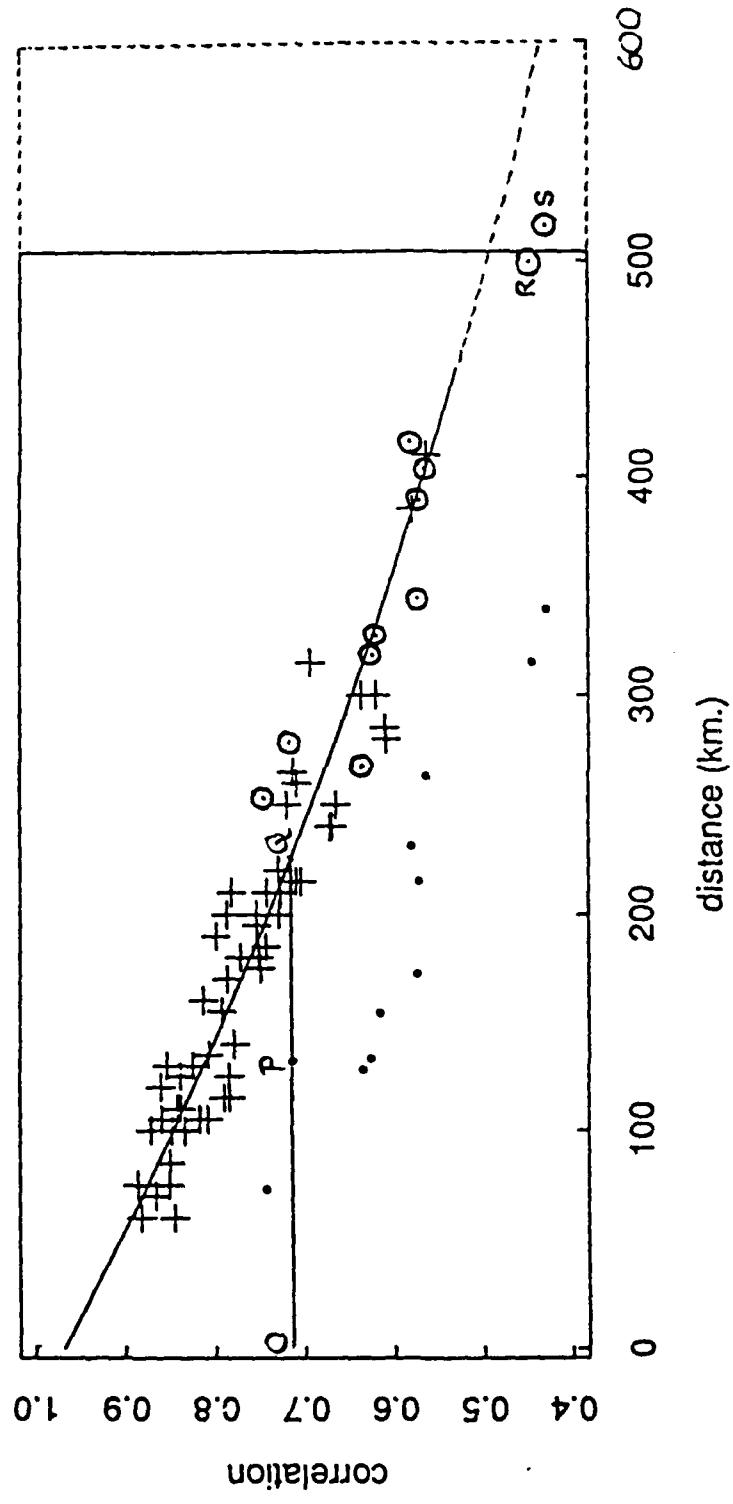
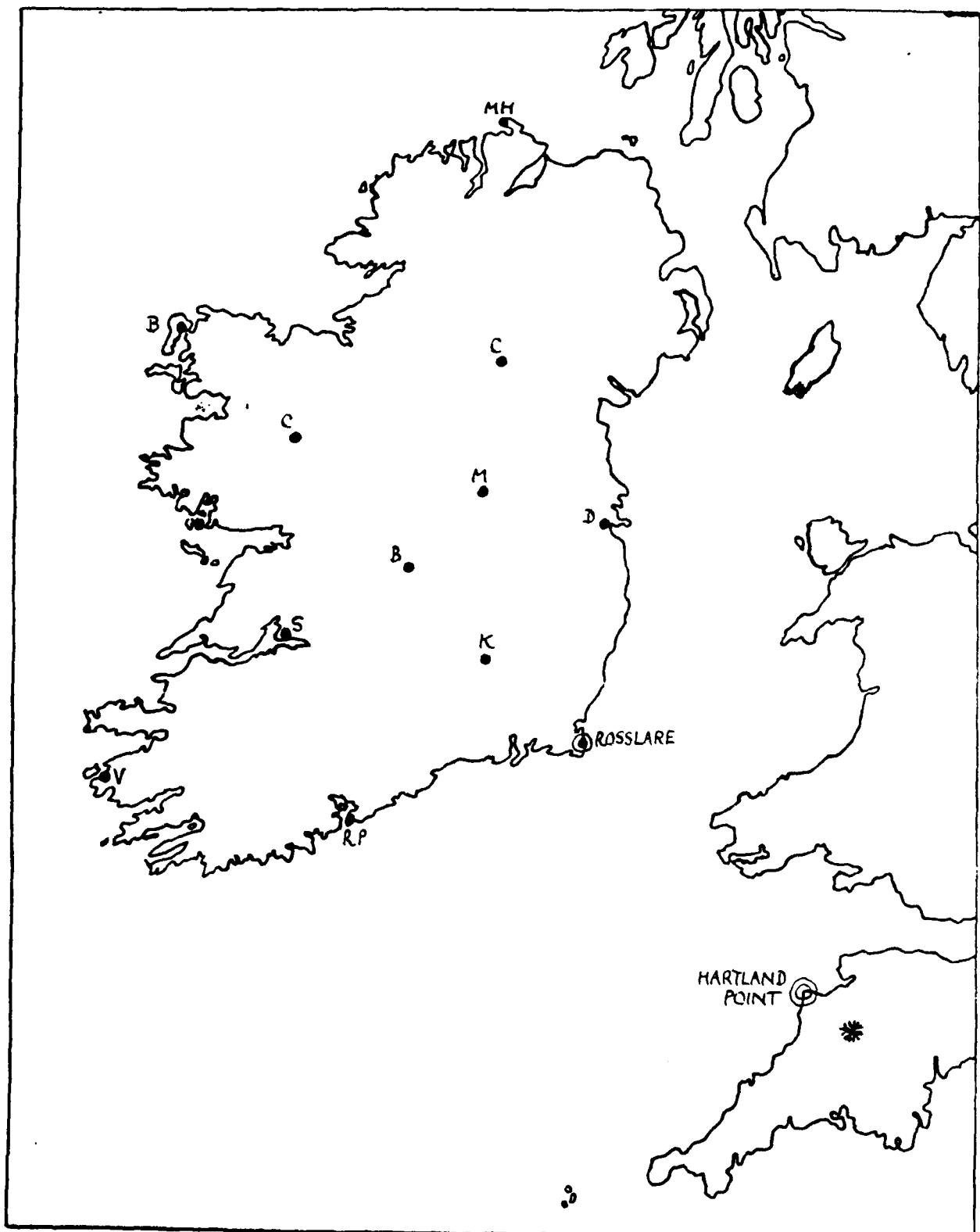


Fig. A. Extension of Fig. 3 of Haslett and Raftery's paper

Fig. D. Extension of Fig. 1 of Haslett  
and Raftery's paper



ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

---

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

1. In Train, Dept of Statistics, Texas A&M University, College Station, TX 77843-3143

2. Name (incl. title)

Dr. John Bertram

3. Affiliation (as you wish it to appear on your printed contribution)

212, Knoxville Society

---

Text of Contribution (Double spaced)

This paper demonstrates once more the importance of long-range dependence for statistical analysis, in particular for the construction of confidence intervals. So far theory and applications were mainly focussed on time series.

Here we have spatial data, though the long-range dependence only occurs in the time dimension. The paper might stimulate research on long-memory processes with a more general index-variable.

The computation of the confidence intervals does not take into account that  $d$  (and also the ARMA-parameters) has to be estimated. Is the effect of estimation negligible ? For instance in the case of the location parameter of a process with a one-dimensional index variable such confidence intervals are clearly too narrow so that the variability of  $d$  has to be build into the procedure. It might be possible to use similar techniques for the model considered in this paper.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

1. Department of Statistics  
La Trobe University  
Bundoora, Vic. 3083. Australia.

2. Name (incl. title)

Dr. J.B. Carlin

3. Affiliation (as you wish it to appear on your printed contribution):

La Trobe University  
Australia.

Text of Contribution (Double spaced)

See attached.

This is an impressive piece of applied statistics. The authors have synthesized several ideas from time series and spatial statistical modelling, in a novel and imaginative way, in order to address a practical problem of considerable difficulty.

A feature of the paper is the use of long-memory time series models. It is salutary to see such clear evidence in these data of the need for models that go beyond the finite spectra of the ARMA class. The analysis presented shows that inferences based on inappropriate short-memory models may be quite misleading when it comes to assessing the variability or uncertainty of estimates of long-term levels. Unfortunately, with shorter time series, it may be much more difficult to assess the nature of low-frequency variation by examining the data (i.e.  $d$  may be hard to estimate). Nevertheless, we should be aware of the potential sensitivity in conclusions to such features of fitted models (Carlin, 1987; Carlin and Dempster, 1988).

On a more technical level, the authors have developed a new and apparently very successful method of approximating the likelihood of the fractionally differenced ARIMA( $p, d, q$ ) process. Further details justifying the method, as well as some systematic evaluations of its performance, would be welcome, as this could be a major contribution towards overcoming the computational difficulties that are a major constraint in the wider application of long-memory models. The computational times quoted by the authors seem consistent with my own experience. Even using the authors' approximation, maximum likelihood estimation seems bound to be computationally costly: it would be interesting to know something of the numerical maximisation algorithm they have used.

Finally, a few comments about the applied problem. The authors' modelling success, as reflected by the almost uncanny agreement of the theoretical and empirical (cross-validators) mean squared errors shown in Table 1, relies on some remarkable empirical regularities observed in their data. For instance,

they argue that it is reasonable to assume a common seasonal pattern, and indeed the same univariate time series structure, for each of their sites, as well as assuming the simple isotropic spatial dependence model (excluding the unfortunate Rosslare). These assumptions could well be violated in countries other than Ireland, with its maritime climate and relatively low relief, so that caution must be exercised in the extension of these methods to other locations. Also, of course, from a limited amount of data at a new, candidate windpower site, it might be difficult to assess whether or not the site has peculiarities like those of Rosslare. Here the input of expert meteorological knowledge would presumably be important. Another feature that weighs heavily in the real-world conclusions of the study is the use of the simple model for expected power output, given by (5.2) and supported by the data of Figure 6. This enables the authors to predict power output simply from an estimate of the long-term mean of the square root of daily wind speed. I wonder if there is any physical rationale for (5.2), or perhaps empirical evidence to support it from other sources? Finally, in Section 5 one might assume that the quantity of ultimate interest,  $\overline{V_{it}^3}$ , should be approximately a *continuous time* average: what is the justification for using the average of *hourly* wind speeds instead?

#### Reference

Carlin, J.B. and Dempster, A.P. (1988) "Sensitivity analysis of seasonal adjustments: Empirical case studies", *J. Amer. Statist. Assoc.*, to appear.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

---

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)	1. Professor N. A. C. Cressie 102E Snedecor Hall, Iowa State University, Ames, IA 50011/U.S.A.
2. Name (incl. title)	Professor N. A. C. Cressie Professor F. Pesarin
3. Affiliation (as you wish it to appear on your printed contribution)	Cressie (Iowa State University) Pesarin (Università degli studi di Padova)

---

Text of Contribution (Double spaced)

See attached sheets

All data have space-time labels, although in many cases it is thought that this information need not be used in the statistical analysis. Drs. Haslett and Raftery have presented us with a study and overwhelming evidence where these labels are very important for forecasting wind speed and energy at unobserved locations. There is a dearth of space-time statistical models in the literature, we think because estimation and distribution theory is difficult for them. The authors have considered a model for which limited inference results are available, and have filled the gaps with cross-validation and conjecture. We congratulate them on their ingenuity and adept handling of a difficult problem.

We have several comments and questions we would like to present for the authors' consideration.

1. We do not believe we can obtain their data set from the published literature; we encourage the authors to make it available for others to perform alternative analyses.
2. Is there any advantage to analyzing power directly, rather than building a model for wind speed and then converting to power?
3. Why did the authors drop two stations, Cork and Casement, from the fourteen reported by Haslett and Kelledy (1979)? They are spatially close to Roche's Point and Dublin, respectively, and would allow verification of the small-lag correlation behaviour assumed in (3.3).
4. Choice of exponential covariance in (3.3) implies sample paths that are continuous (when there is no nugget effect) but not differentiable. At the scale of spacing of the synoptic stations, this does not matter, but if wind turbines were to be clustered around centers of population, small-scale sample path behaviour is important. If the fitted space-

time model were used to simulate the wind speed at all scales, the answers may be inappropriate for certain questions at the small scale. The rate of approach of the spatial correlation function to the abscissa could be checked by using data from Cork and Casement, two synoptic stations omitted by the authors.

5. We see a spatial inhomogeneity in the time series of Figure 4. Stations Valentia, Roche's Point, and Rosslare do not seem to have the same long-range dependence as the other stations. Was this seen in the diagnostics used on the residuals from the authors' model (4.1)(which assumes a temporal operator on spatially stationary errors that is homogeneous across space)?
6. Residual's are different from errors; residuals contain spurious correlations that bias estimation of the error correlation structure. In fact the authors' "original" data are residuals, having first been deseasonalized.
7. The seasonal component was assumed deterministic for all the calculations, but clearly it is estimated.
8. The authors make the point that under long-range temporal dependence, there is little loss of asymptotic efficiency in using unweighted means. A similar phenomenon occurs in space; Kramer and Donninger (1987) give a result of this type for a simultaneous spatial autoregressive Gaussian process.
9. The wind-speed data exhibit high spatial correlation, severely reducing the effective number of "spatial observations." Without the spatial homogeneity assumption referred to (and questioned) in 5., estimators would be highly variable.

10. We think the term "kriging estimator" is inappropriate. Kriging refers to prediction, which we think should be distinguished from estimation. We believe that kriging is what is needed here, but that estimation ignores the question of variability in the potential observations. Data are recorded using instruments that will be different from the turbines that will actually generate the power. Thus it is the variability with regard to the turbines that should be considered. This is known as the "change of support problem" in the geostatistics literature, and is ignored by considering inference on means.

Additional Reference:

Kramer, W. and Donninger, C. (1987). Spatial autocorrelation among errors and the relative efficiency of ols in the linear regression model. Journal of the American Statistical Association, 82, 577-589.

May 25, 1988

Contribution to the Discussion of "Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource" by Haslett and Raftery.

A. P. Dempster, Harvard University

The paper is interesting and authoritative, and quite remarkable for the wide range of issues considered in so brief a report, including exciting new methodology for a problem of major economic importance. My comments are limited to matters pertaining to statistical modelling, and are based on experience with similar time series models also estimated by m.l., albeit only univariate and much shorter series. Readers may find a forthcoming paper by Carlin and Dempster (1988) more accessible than the paper by Carlin, Dempster, and Jonas, and the Carlin thesis, as cited.

A basic difficulty in dealing with 11 simultaneous and long ( $n = 6574$ ) time series is the possible wild proliferation of parameters. The authors deal with this by ruthlessly enforcing parsimony, eg, using common fixed seasonal patterns, and common simple whitening filters, for all the series. The simple linear model for space correlation implicitly assumes that the pairwise cross-spectra are constant across frequency and all have zero phase shifts. While the extreme parsimony renders m.l. feasible, I wonder if it is not overdone, especially with such long series. In particular, I wonder if data analysis could show dependence of correlation on frequency and perhaps location-related phase shifts at different frequencies.

My main comment is to question the authors' approach to long-memory dependence. It seems to me that the Fig. 5 periodograms of AR(9) whitened series removes not only "short-memory" dependence, but in fact makes the spectra flat across 99% of the frequency range, ie, from .005 to .5, and shows only a hint of increase across a further .8%, ie, from .001 to .005. Thus only about 1/500 of the periodogram ordinates suggest further long-memory dependence, and sampling theory for these few points is not yet well understood, so they are hard to interpret, leading me to question whether  $d$  can be safely estimated from the AR(9) residuals. In addition, the AR(9) itself is quite capable of representing something indistinguishable from long-memory dependence via roots near unity.

A different criticism applies to the m.l. procedure, and applies also to my own work with Carlin. The high apparent accuracy with which  $d$  is estimated results from, in effect, using the fractional differencing term in the model to shape spectra across the full frequency range 0 to .5. A very different value of  $d$  could be operating near 0 frequency, yet the procedure could completely miss this fact. Indeed, the low frequency power need not be a power law at all. For example, it might have peaks near the 11 or 22 year sunspot cycles, yet the data would have no sensitivity. It is sobering that with so much data we really cannot identify important low frequency phenomena without strong assumptions. What are the practical implications for forecasting energy yields?

#### ADDITIONAL REFERENCE

Carlin, J. B. and Dempster, A. P. (1988) Sensitivity analysis of seasonal adjustments: empirical case studies. To appear. J. Amer. Statist. Assoc. 83.

Comment on "Space-Time Modelling with Long-Memory Dependence:  
Assessing Ireland's Wind Power Resource" by John Haslett and Adrian E. Raftery

Peter Guttorp and Paul D. Sampson

Department of Statistics  
University of Washington  
Seattle, WA 98195  
U.S.A.

A unique feature of this paper is the explicit recognition of the dependence of spatial correlation on temporal scale in this application. The resulting definition and interpretation of spatial correlation is intrinsically different from that used when there is no time replication (common in many geostatistical spatial studies), or when data are time-averaged. We raise two questions and propose an alternative, nonparametric approach to Haslett and Raftery's (H&R) spatial covariance model. This approach does not require a stationary or isotropic covariance structure, and so obviates the *ad hoc* approach of eliminating Rosslare from the analysis.

The long-term memory evidence is convincing. However, the authors do not suggest any explanation of it. Can it be related to climatological principles? Similarly, can meteorological theory be used to model the seasonal variation? This would seem more appropriate than fitting harmonics. From a data analytic point of view, one may want to use a local smoother with a higher degree of flexibility to estimate the seasonal term. The effect on the spectral estimates of a local smoother is less clear than that of harmonics. Perhaps some insight can be had using Mallow's (1980) concept of linear parts of non-linear smoothers.

In connection with an assessment of solar power potential in British Columbia, we are developing a method for estimating non-stationary anisotropic spatial covariances from repeated observations at a set of stations (Sampson 1986). The solar energy field must be estimated everywhere, not only where short runs of pilot data are available. Since the estimator (3.4) does not apply for extrapolation to a location without pilot data, this requires a spatial analysis more closely related to standard kriging methods. We model *spatial dispersions*  $v_{ij} = \text{Var}(X_i - X_j)$  as a general function of the geographic locations of stations  $i$  and  $j$ , not simply as a function of the distance  $d_{ij}$  between the stations. This is accomplished by applying multi-dimensional scaling (MDS) to the matrix  $(v_{ij})$ , considered as dissimilarities, to obtain a new two-dimensional representation of the sampling stations in which the spatial dispersion function (or variogram) satisfies the common assumption of stationarity and isotropy (*i.e.*, being determined only by metric distances between station locations). Station pairs that are weakly correlated (have large  $v_{ij}$ ) will be located relatively further apart in the MDS representation than they are geographically. We estimate the spatial dispersion  $v_{ij}$  (and thereby the spatial covariance) between any two locations in the geographic plane using the composite of: (a) the monotone relationship between spatial dispersion and the inter-station distances in the MDS representation, and (b) a smooth mapping (computed using thin-plate splines) between the geographic and MDS representations. This mapping embodies the nature of the manifest anisotropy and non-stationarity; it can be depicted graphically using biorthogonal grids (Bookstein 1978).

Applying MDS to the sample covariance matrix for the Irish wind power data (provided to us by Professor Raftery), we obtained Fig. 1. Compare this with the geographic map in Fig. 1 of H&R. The stations around the coast are located relatively further from the stations in the middle of the island, indicating that covariance between coastal stations and inland stations is weaker than that among inland stations. Rosslare is furthest displaced in accordance with its relatively weak covariance with all other stations. Fig. 2 displays the success of MDS in representing the dispersions  $v_{ij}$  as a function of distance in Fig. 1. The authors refer to some studies of robustness to misspecification of the spatial covariance structure. However, these are limited to misspecification of stationary structures. Part of

the non-stationarity in these data is due to a gradient in the station variances: decreasing variance from the northwest to the southeast. Fig. 3 of H&R, a plot only of correlations, does not show this.

Our approach to spatial covariance cannot be directly integrated into the likelihood estimation framework of Section 4. However, H&R's maximum likelihood estimates of parameters of the isotropic spatial correlation function in (4.1)-(4.2) are, in fact, little changed from the preliminary estimates obtained by regressing  $\log\{\text{Corr}(X_{it}, X_{jt})\}$  on  $d_{ij}$ . This suggests that one may simplify the estimation procedure described in section 4 by removing the parameters of the spatial covariance process from the likelihood (*i.e.*, holding them fixed). Then the likelihood is expressed in terms of a fixed estimate of the spatial correlation matrix,  $R$ , for which we would propose substituting our nonparametric estimate of spatial covariance. This estimate could be refined as necessary upon examination of the  $e_i$  in the model checking phase (section 4.4).

*References:*

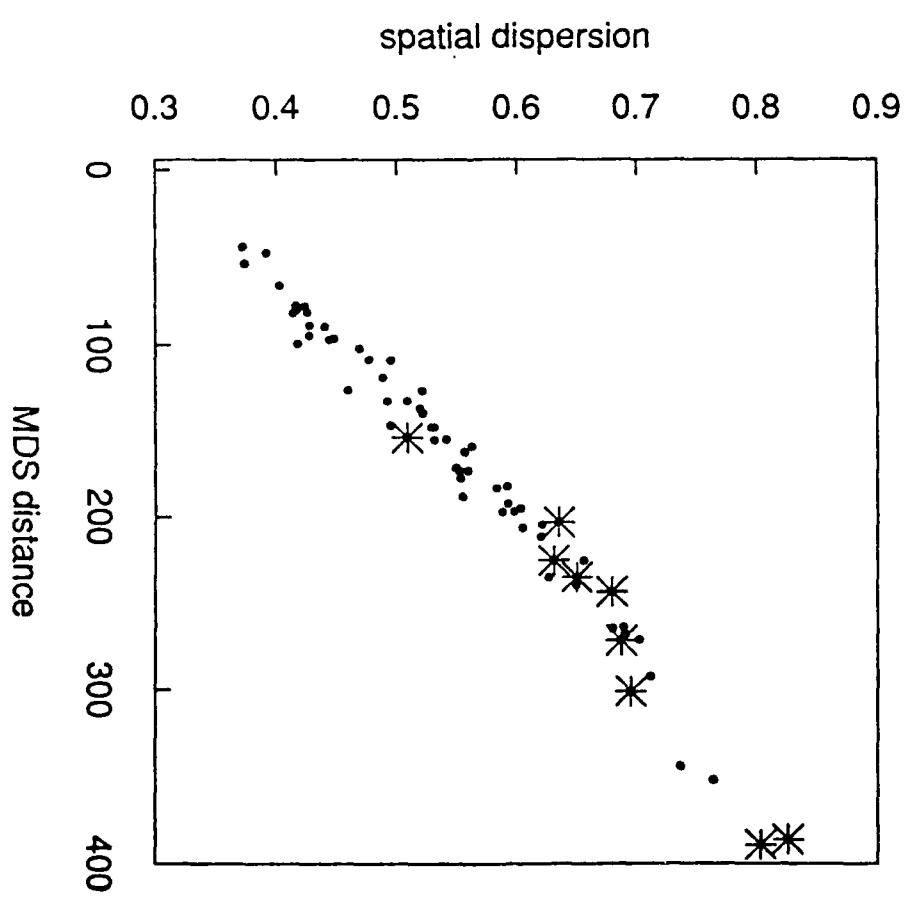
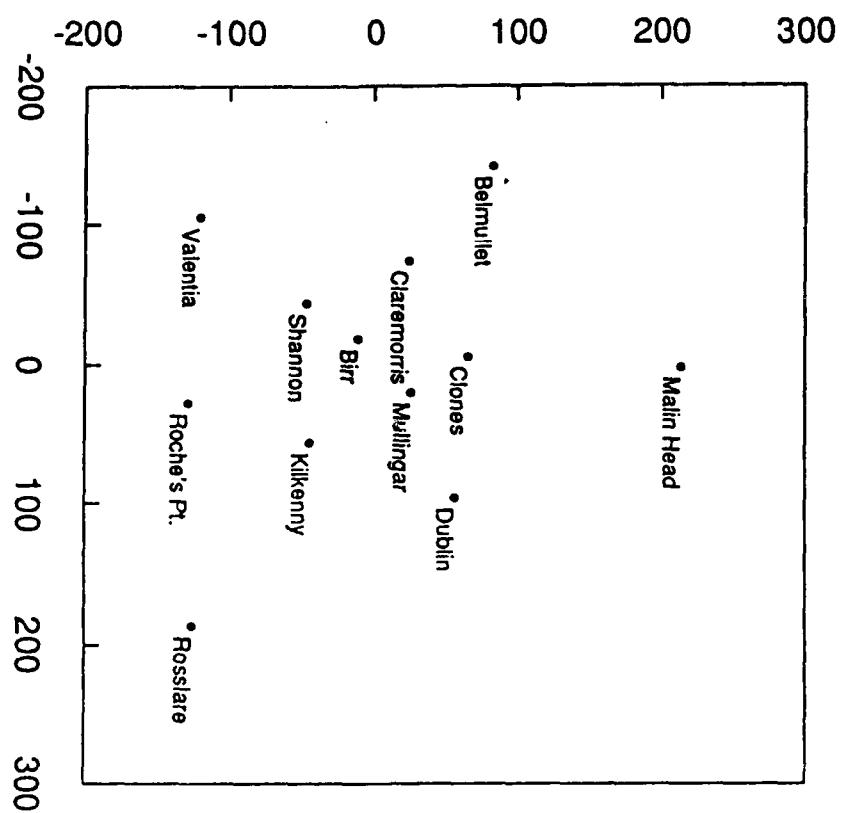
Bookstein, F. L. (1978): *The Measurement of Biological Shape and Shape Change*. Lec. Notes Biomath. 24. New York: Springer.

Mallows, C. L. (1980): Some theory of nonlinear smoothers. *Ann. Statist.* 8: 694-715.

Sampson, P. D. (1986): Spatial covariance estimation by scaled-metric scaling and biorthogonal grids. SIMS Tech. Rpt. No. 102, Univ. of British Columbia.

Fig. 1. MDS representation of the IMS monitoring stations based on estimated spatial dispersions  $v_{ij}$ .

Fig. 2. Plot of spatial dispersion  $v_{ij}$  versus inter-station distance in the MDS representation. Asterisks correspond to station pairs involving Rosslare.



Contribution to the discussion of Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource, by John Haslett and Adrian Raftery, 25th May 1988

In a large applied project such as this there are always alternative approaches possible. Two occur to me.

First, the modeling of the series does not discuss the lagged cross-correlations. Given that the stations are several hundred kilometers apart and that weather patterns tend to move from west to east, I would have expected a delay of up to 12 hours between the west coast and east coast stations. This could be readily modeled using for example the spectral methods of Hannan and Thompson (1974).

Second, it is clear from the periodograms in Figure 4 that the temporal persistence referred to is on a time scale of several years. (It could not be much less since the seasonal component has been removed and the AR(9) model would remove most of the variance over shorter periods.) In my experience with long term meteorological data such temporal persistence is likely to be due in part to changes in the measuring equipment and in the environment around the measuring station rather than in the weather. It is not unusual for stations themselves to be moved. However the methods of this paper could be used to predict the daily velocity measures at each station from the measures at the other stations and the discrepancy between the actual and predicted records could be expected to highlight sudden changes in the mean. This can then be corrected if felt justified. It is likely that there remain a long-memory dependence component but on a reduced scale.

Hannan, E.J. and Thompson, P.J. (1971) The estimation of coherence and group delay, Biometrika, 58, 469-481.

Dr John Henstridge  
Perth, Western Australia

Dr D A Jones (Institute of Hydrology); Given the contrast in performance between short and long-memory models, it would be interesting to include medium-memory models for consideration. Such models might reasonably be defined in terms of their partial autocorrelations. For example, for a model with three parameters a, b and c, let

$$\phi_{11} = a, \phi_{22} = b, \phi_{jj} = c \quad (3 \leq j \leq M), \phi_{jj} = 0 \quad (M < j),$$

where  $M=50$  or  $100$ . This of course corresponds to an  $AR(M)$  process. An alternative model might allow  $\phi_{jj}$  to taper linearly to zero, but sample estimates might suggest more appropriate behaviour.

Some of the difficulties reported with  $ARIMA(p,d,q)$  processes arise from the calculation of their partial autocorrelation functions: one possibility is to move to models parameterised directly via these functions, much as above, with a suitable behaviour for  $\phi_{jj}$  as  $j$  increases. Modelling directly in terms of the partial autocorrelations would fit in with the authors' existing estimation scheme, while avoiding the need for approximations. The only disadvantage seems to be that the rather mesmeric statements of model structure, such as equation (4.1), are lost.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

2. Name (incl. title)

3. Affiliation (as you wish it to appear on your printed contribution)

1. Richard W. Katz \_\_\_\_\_  
\_\_\_\_ NCAR/ESIG \_\_\_\_\_  
\_\_\_\_ P.O. Box 3000 \_\_\_\_\_  
Boulder, CO 80307 U.S.A.

\_\_\_\_ Dr. Richard W. Katz \_\_\_\_\_  
\_\_\_\_ National Center for Atmospheric \_\_\_\_\_  
Research

**Text of Contribution (Double spaced)**

-----attached-----

This paper provides a useful method for synthesizing several statistical characteristics that are typical of climatic variables such as wind speed. These characteristics include non-normal distribution, seasonal cycles, and temporal and spatial correlation. The most novel aspect of this work concerns the issue of long-memory dependence. Models that possess long-memory dependence are sometimes considered in the water resources literature, especially as one possible chance mechanism to explain the origin of the so-called "Hurst phenomenon" (Hosking, 1984). However, such models are not routinely considered by climatologists in fitting variables such as wind speed.

Convincing evidence is provided in this paper that taking into account temporal correlation (both short-memory and long-memory) is necessary for providing reliable standard errors in the estimation of mean wind speed. It should be noted that climatologists are well aware of the need to correct for the effect of short-memory correlation on the standard error of time averages. In particular, a formula that is essentially a special case of (4.10), but ignores long-memory correlation, has been frequently employed in the meteorological literature (e.g., Jones, 1975).

Finally, stationarity on an interannual time scale has been assumed in all of the analyses contained in this paper. But one of the issues in climatology over which the most controversy currently exists concerns whether or not the climate is undergoing permanent change (e.g., Wigley and Jones, 1981). Moreover, nonstationarity is an alternative chance mechanism to long-memory dependence for explaining the Hurst phenomenon (Bhattacharya et

al. 1983). Consequently, the conclusions of this paper relating to the efficient allocation of resources for measuring wind speed need to be qualified.

#### REFERENCES

Bhattacharya, R.N., Gupta, V.K. and Waymire, E.C. (1983) The Hurst effect under trends. J. Appl. Prob., 20, 649-662.

Hosking, J.R.M. (1984) Modeling persistence in hydrological time series using fractional differencing. Water Resources Research, 20, 1898-1908.

Jones, R.H. (1975) Estimating the variance of time averages. J. Appl. Meteor., 14, 159-163.

Wigley, T.M.L. and Jones, P.D. (1981) Detecting  $CO_2$ -induced climatic change. Nature, 292, 205-208.

ROYAL STATISTICAL SOCIETY  
25 Entford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

---

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)	1. <u>Seminar für Statistik</u> ----- ----- CH-8092 Zurich, Switzerland
2. Name (incl. title)	Professcr Hans R. Künsch ----- ----- ETH Zurich ----- -----
3. Affiliation (as you wish it to appear on your printed contribution)	

---

Text of Contribution (Double spaced)

I was very pleased to see here another example of data which clearly exhibit long-range dependence. It is the first multivariate example I know of. The model considered by the authors is a simple and useful subclass among the large number of possible multivariate models. It implies that not only all autocovariances and autospectra, but also all crosscovariances and crossspectra are proportional. I guess that the authors have checked this assumption at an exploratory stage.

The approximation to log likelihood studied by Fox and Taqqu (1986) and Beran (1986) is Whittle's approximation. It is available also in the multivariate case, see Whittle (1953, Th. 6). For the model (4.1) it equals

$$\log\sigma_\epsilon^2 + \log \det R + \sigma_\epsilon^{-2} \int |1 - e^{i\lambda}|^{2d} |\phi(e^{i\lambda})|^2 |\theta(e^{i\lambda})|^{-2} \sum_{j,k} (R^{-1})_{jk} I_{N,jk}(\lambda) d\lambda$$

where  $I_{N,jk}$  is the crossperiodogram. Approximating the integral by a sum an evaluation of this expression should not take much CPU-time.

Finally I would like to propose a slight variant of the estimator (3.4) and its approximate variance (4.10). For simplicity we take in the estimation problem of Section 3  $N = Mn$  and  $t_0 = N - n + 1$ . Other values of  $t_0$  can be handled similarly. We consider the following estimator depending on coefficients  $\alpha_j$ :

$$\hat{\mu}_k = n^{-1} \sum_{t=t_0}^N X_{kt} + \sum_{j \neq k} \alpha_j (n^{-1} \sum_{t=t_0}^N X_{jt} - N^{-1} \sum_{t=1}^N X_{jt})$$

Under the model (4.1) the covariance between block sums  $\sum_{t=1}^n X_{it}$  and  $\sum_{t=sn+1}^{(s+1)n} X_{jt}$  is for large  $n$  approximately

$$\sigma_e^2 r_{ij} c(\phi, \theta, d) n^{1+2d} \frac{1}{2} (|s+1|^{1+2d} - 2|s|^{1+2d} + |s-1|^{1+2d}),$$

see Cox (1984). If these covariances hold exactly, the optimal coefficients  $\alpha_j$  can be obtained easily. The variance of  $\hat{\mu}_k$  is then equal to

$$\sigma_e^2 c(\phi, \theta, d) n^{2d-1} (1 - u_M^2/v_M (1 - a_{kk}^{-1}))$$

where  $u_M = 1 - M^{-1} + M^{-1}(M-1)^{2d+1} - M^{2d}$ ,  $v_M = 2u_M - 1 + M^{2d-1}$ ,  $a_{kk} = (R^{-1})_{kk}$ . The factor  $1 - u_M^2/v_M (1 - a_{kk}^{-1})$  gives the decrease of the variance due to the information at other sites. Because  $u_M$  and  $v_M$  converge to one rather slowly, it can be close to one even if  $a_{kk}^{-1}$  is small, i.e. the spatial dependence is strong. This shows that the information from other sites is useful only if the records there are much longer than at the site of interest. The statement of the last paragraph of the paper thus seems too optimistic to me.

#### Additional References:

Cox,D.R.(1984) Long-range dependence: a review. In Statistics: An Appraisal (H.A. David and H.T. David, eds.), pp. 55-74, Iowa State Univ. Press.

Whittle, P.(1953) The analysis of multiple stationary time series. J. Royal Statist. Soc. B, 15, 125-139.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

1. Department of Statistics,  
University of Hong Kong,  
Pokfulam Road, Hong Kong.

2. Name (incl. title) Dr. W.K. Li

3. Affiliation (as you wish it to appear on your printed contribution)

University of Hong Kong

Text of Contribution (Double spaced)

The authors are to be congratulated for their interesting work in generalizing the fractional time series process to the space-time situation.

I would like to concentrate my comments on the modelling aspect. In practice, it seems rather unlikely that all  $m$  stations exhibit the same long term and short term autocorrelation structure. Therefore model (4.1) appears to be a simplification and a more general model with  $d$ ,  $\Phi(B)$  and  $\Theta(B)$  depending on  $i$  could be entertained. Of course, the modelling would become more difficult. In a recent report, Hui and Li (1988) consider fractionally differenced periodic processes where  $d$  or  $\Phi(B)$  are allowed to vary over different seasonal periods. The results

...../may

may be applicable to the present problem. Since model (4.1) only makes use of the information provided by the distances between stations it is more akin to the so called contemporaneous ARMA models studied by Camacho, McLeod and Hipel (1987) than to a spatial time series over a rectangular lattice. Thus the approach of Mardia and Marshall (1984) may not be needed here. It seems also to me that some sort of approximations to  $\nabla^d$  or the exact likelihood is unavoidable in practice and in my experience such approximations do appear to be rather satisfactory with sufficiently long records of data. Finally, the maximum likelihood estimate  $\hat{\alpha}$  is rather close to one although its approximate standard error is only 0.0013. Have the authors considered a model with  $\alpha$  set equal to one?

Camacho, F., McLeod, A.I. and Hipel, K.W. (1987). Contemporaneous Bivariate Time Series. *Biometrika*, 74, pp.103-13.

Hui, Y.V. and Li, W.K. (1988). On Fractionally Differenced Periodic Processes. Manuscript, Chinese University of Hong Kong and University of Hong Kong.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

---

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)
2. Name (incl. title)
3. Affiliation (as you wish it to appear on your printed contribution)

---

1. BROWN UNIVERSITY, Box F  
PROVIDENCE, RI, 02912  
USA

---

ALAN LIPPMAN, Vist. Asst. Prof  
BROWN UNIVERSITY  
DIVISION OF APPLIED MATHEMAT

---

Text of Contribution (Double spaced)

Raftery and Haslett have proposed a reasonable model for daily average wind speed in Ireland. The clean spatial correlation structure implied by figure 3, enables the authors to make effective use of a kriging type estimator for the expected daily mean wind speed, which yields, at any location, good estimates based on little data. The model they propose for the daily mean provides remarkably reliable estimates of the variance of the kriging estimate of the expected daily mean.

It is unfortunate, though understandable, that the authors could see no way to estimate the distribution of the wind speed (not the daily mean). If one had the true distribution of wind speeds it would be trivial to calculate the expected power production, as power production is a known, turbine dependent, nonlinear function of wind speed.

The authors instead use a clever two-part approach to achieve their goal, first modeling the daily mean and then using the model to estimate expected power production. It is upon the second part, involving the use of the kriging estimate and its error bounds, that I would like to comment.

While I am not well versed in the mechanics of turbines, the authors' assumption that power production is proportional to the power in the wind appears hazardous to me, as this ignores the effects of extrema. This is a point the authors mention briefly, but could prove important. Turbines shut down at high wind speeds. Ignoring this could lead to over-estimating power production. I assume the authors have already considered this, but I would be interested to see a modified figure 6, plotting log power produced (for a specific type of turbine) versus log daily mean.

Granting that power production is proportional to the power in the wind, I wonder if an improvement could not be made in its estimation by using more than just the kriging estimate of the daily mean and its error bounds. It should be possible at a new site to estimate some statistics of the wind speed, for example the variance of the square root wind speed. I pick this quantity since the authors observed that the square root wind speed was approximately normal. An estimate of this variance, when used in conjunction with an estimate of the expected daily mean might yield a better estimate of the expected cubed wind speed. A 20 day sample period yields 480 hourly samples, enough, perhaps, for a reasonable estimate of this variance, and while there would be seasonal effects to consider, I would not anticipate anything like long-memory dependence. So, another modification of figure 6, this time by adding a third dimension, variance of the square root wind speed, might be revealing.

I would like to thank the authors for a thought provoking paper, and a pleasing example of the application of spatial statistics to a difficult real-world problem.

Discussion to the paper by 'Haslett and Raftery' on 25th May 1988.

Professor K. V. Mardia (University of Leeds): First of all, let me congratulate the authors for a very stimulating paper. The terminology of "kriging" estimation" in the paper could be somewhat misleading. Usually kriging is used for prediction whereas in the paper the term is used for parameter estimation. In fact, let  $X = (X_1, X_2)'$  be  $N(\mu, \Sigma)$  with the usual partitioning for  $\mu$  and  $\Sigma$  where  $X_2$  is the scalar variable at the new site. Then, from conditional expectation we have

$$\mu_2 = E(X_2 | X_1) + \Sigma_{21} \Sigma_{11}^{-1} (X_1 - \mu_1).$$

Their estimator  $\hat{\mu}_2$  of  $\mu_2$  at the new site, given by Eq. (3.4), is obtained on replacing in the R.H.S. of the above equation,  $\mu_1$  by the sample mean of all the  $N$  observations, and  $E(X_2 | X_1)$  and  $X_1$  by the sample means of  $X_2$  and  $X_1$  based on the  $n$  observations respectively,  $n < N$ . Of course, the tools in both cases are similar as one is using (a) the conditional expectation and (b) a covariance scheme.

I do not believe that the robustness of  $\hat{\mu}_2$  for values of  $\alpha$  and  $\beta$  follows from the previous studies related to prediction. However, one might expect it to be true. But as it has been pointed out by the authors, the variance of  $\hat{\mu}_2$  will be definitely influenced by the estimated values of  $\sigma_X^2$ ,  $\alpha$  and  $\beta$ . Therefore, an efficient method of estimation is desirable. It is common in Geostatistics to plot semivariograms rather than correlation functions, particularly for processes which have stationary increments but are not stationary. Might not the use of semivariograms also be fruitful for long range correlations?

The authors indicate that combining known results on asymptotic normality of Mardia and Marshall (1984) with others, they could obtain similar results for their model. However, the nugget parameter causes some theoretical difficulty as it lies on the boundary of the parameter space. For a further

discussion of this topic see Watkins (1988).

The authors removed the data at Rosslare in estimating  $\alpha$  and  $\beta$ . This might indicate that there is some effect of the wind-direction in general. The behaviour of "co-kriging estimation" through wind velocity rather than just wind speed will depend heavily on the underlying cross-covariance structure. Which cross-covariance scheme was used by the authors?

#### References

Watkins, A.J. (1988) Some Aspects of Statistical Inference in the Spatial Linear Model. Ph.D. Thesis, Department of Statistics, University of Leeds.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

---

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

1. Dept. of Statistics  
University of Western Ontario  
London, Ontario CANADA N6A 5B9

2. Name (incl. title) Dr. A.I. McLeod

3. Affiliation (as you wish it to appear on your printed contribution) University of Western Ontario

---

Text of Contribution (Double spaced)

Contrary to a statement made at the beginning of the second last paragraph of Section 4.3, the asymptotic distribution of the parameter estimates in a univariate ARIMA  $(p, d, q)$  with  $|d| < 0.5$  has been derived by Li and McLeod (1986).

The model used by Haslett and Raftery can be viewed as a long-memory extension of the CARMA (contemporaneous ARMA) model of Camacho, McLeod and Hipel (1987 a,b).

REFERENCES

Camacho, F., McLeod, A.I., and Hipel, K.W. (1987a). Contemporaneous bivariate time series. *Biometrika*, 74, 103-113.

Camacho, F., McLeod, A.I., and Hipel, K.W. (1987b). Multivariate contemporaneous ARMA models with hydrological applications. *Stochastic hydrology applications*, 1, 141-154.

Li, W.K and McLeod, A.I. (1986). Fractional time series modelling. *Biometrika*, 73, 217-21.

# Discussion of Haslett & Raftery <sup>1</sup>

Yosihiko OGATA

The Institute of Statistical Mathematics,  
Minami-Azabu 4-6-7, Minato-ku, Tokyo 106

It is my great pleasure to comment on the very stimulating paper by Drs John Haslett and Adrian Raftery. I am concerned in the fact suggested from Figure 5. That is to say, all periodograms in this figure have common peaks at the one year period, in spite of deseasonalisation of the data using the estimate in Figure 2. This indicates that the seasonal effect at each station may not be quite the similar to those at the other stations. In this occasion, I would like to describe a possible analysis for such case in relation to the interpolation problem.

Consider the original data  $X_{it}$  as the spatio-temporal data  $X(t_i, x_i, y_i)$  on  $[0, T] \times A$ , where  $A$  is the rectangular region of Figure 1 including Ireland. Then consider a three dimensional spline function  $h(t, x, y | c)$  parameterized by  $c$ . Since quite many number of parameters will be required to get the sensible estimates of the trend, I consider the penalized log likelihood, where, besides the standard roughness penalties for the spline function  $\Phi_1(h) = \int_A \int_0^T \{ \frac{\partial^2 h}{\partial t^2} \}^2 dt dx dy$ , and  $\Phi_2(h) = \int_A \int_0^T \{ (\frac{\partial^2 h}{\partial x^2})^2 + 2(\frac{\partial^2 h}{\partial x \partial y})^2 + (\frac{\partial^2 h}{\partial y^2})^2 \}^2 dt dx dy$ , the seasonality constraint

<sup>1</sup>paper read at the RSS meeting 25 May, 1988

is given by  $\Phi_3(h) = \int_A \int_{T_0}^T \{h(t - T_0, x, y) - h(t, x, y)\}^2 dt dx dy$ , where  $T_0 = 365.24$  days. Or, alternatively, we may regard the original data as the superposed spatio-temporal data  $X(s_i, x_i, y_i)$  on  $S \times A$ , where  $S$  is the one-dimensional torus being identical to  $[0, T_0]$ , and a very heavy weight is imposed to the penalty for the periodicity,  $\Phi_3(h) = \int_A \{h(0, x, y) - h(T_0, x, y)\}^2 + \{\frac{\partial h}{\partial t}(0, x, y) - \frac{\partial h}{\partial t}(T_0, x, y)\}^2 + \{\frac{\partial^2 h}{\partial t^2}(0, x, y) - \frac{\partial^2 h}{\partial t^2}(T_0, x, y)\}^2 dx dy$ .

To obtain the suitable weights, I employ the Bayesian interpretation of the penalized likelihood (Akaike, 1979): The sum of the weighted penalties are considered to be proportionate to the logarithm of prior probability density  $\pi(\mathbf{c} | w_1, w_2, w_3)$  of the parameters  $\mathbf{c}$ , and the penalized log likelihood is considered to be the log posterior distribution. Then the marginal of the posterior (the Bayesian likelihood),  $\Lambda(\sigma, w_1, w_2, w_3) = \int L(\mathbf{c} | \sigma) \pi(\mathbf{c} | w_1, w_2, w_3) d\mathbf{c}$ , is maximized to obtain the optimal weights.

The estimated spline function can be used for interpolating the seasonal effect at any locations. Further the so-called universal kriging procedure, subtracting the trend of the estimated spline, can then be carried out. On the other hand, assuming that the sample space of the spatio-temporal random field are restricted to a class of smooth spline functions, we have an alternative kriging method using the Gaussian posterior distribution of the parameter  $\mathbf{c}$ . See Ogata (1988) for the longer version of the present comments, and also Ogata and Katsuma (1988) for some details and numerical performance for the related spatial problems.

## References

Akaike, H. (1979). Likelihood and Bayes procedure, *Bayesian Statistics*, (eds. J.M. Bernard et al.) University Press, Valencia, Spain.

Ogata, Y. (1988). Comment on "Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resources" by John Haslett and Adrian Raftery, *Research Memorandum*, No. 351, 1-5.

Ogata, Y. and Katsura, K. (1988). Likelihood analysis of spatial inhomogeneity for marked point patterns, *Ann. Inst. Statist. Math.* **40**, 29-40.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)
2. Name (incl. title)
3. Affiliation (as you wish it to appear on your printed contribution)

1. Dept. of Statistics  
Birkbeck College  
Malet Street WC1E 7HY

DR. W. D. RAY  
(Birkbeck College, London)

Text of Contribution (Double spaced)

Dr W D Ray (Birkbeck College, London)

Current statistical literature not infrequently deals with a far too idealistic model which is deemed to be sacrosanct, a theory is then developed to the finest detail with the pious hope that sometime, somewhere data will be found to fit. It is nice to see a paper which is more data-orientated, and which checks out early features through exploratory analysis to judge, for example, likely transformations, and levels of aggregation. Another time series paper by Harvey and Durbin two years ago on seat belt legislation was also in this vein, but such contributions are not as common as they ought to be.

The paper is fairly self-contained and complete, but I have a few peripheral comments. I was surprised that the estimated seasonal effect in Fig 2 required several harmonics, the scatter seems to indicate that fewer would have sufficed. The striking homogeneous short-memory

autocorrelations of Fig 4 are remarkable, particularly the positive aspects. So too is the common pattern of low frequency-long memory persistence in Fig 5. Hence the need for fractional differencing, and this data provides a good example of it's necessity.

The commonality feature of the wind data at the synoptic sites in Ireland is fortunate to allow the relative simplicity of model 4.1 and 4.2, but this feature may not be present in other applications when some clustering may be necessary.

It was not too surprising that the numerical aspects of maximum likelihood estimation are a problem here, a factor which also becomes acute when handling non-linear time series with large data sets. Thus the approaches to obtain approximations are to be commended.

The comment in 4.4 that non-linearities were not present in this wind data could have been amplified by providing a few statistics, which could then have been useful for future researches. The agreement of the M.S.E.'s from 4.10 with the empirical results seem rather flattering to the approximation.

This work is a very good example of time series modelling carried out in the true spirit of data leading the way. The class of models, 4.1 and 4.2 are wide enough to be of use in a greater variety of applications, and probably will.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

1. DEPT OF STATISTICS

UNIVERSITY OF GLASGOW

GLASGOW, G12 8Q0

2. Name (incl. title)

DR E. M. SCOTT

3. Affiliation (as you wish it to appear on your printed contribution)

UNIVERSITY OF GLASGOW

Text of Contribution (Double spaced)

I would like to congratulate the authors on a very interesting paper and to make some comments on related problems.

- (1) My first comment concerns the non-Bayesian nature of the analysis. Given the nature of the problem (and others in environmental sciences), it would seem likely that prior information on a specific site would be available and that potential covariates might exist, which could and should be incorporated in the analysis.
- (2) Secondly, an important problem, not tackled in the paper, would involve the question of the siting of the synoptic stations, and whether there might be any possibility of developing the modelling approach to identify "optimal" sites for wind farms, which could then be investigated in more detail.
- (3) Finally, the removal of the 12th station from the analysis raises interesting questions concerning the coarseness of the synoptic site grid relative to the degree of spatial variability in wind over a large geographical area.

There must be many sites where the global wind model is difficult to apply due to local conditions.

How should one balance siting and number of synoptic stations with the spatial variability of the response?

ROYAL STATISTICAL SOCIETY  
15 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

1. Department of Statistics  
University of Chicago  
5734 University Ave.  
Chicago, IL 60637 USA

2. Name (incl. title)

Professor Michael Stein

3. Affiliation (as you wish it to appear on your printed contribution)

University of Chicago

Text of Contribution (Double spaced)

Standard asymptotic results often do not apply in a spatial context. For example, in Section 3, the authors state that the estimate  $\hat{\mu}_k$  will be "approximately normally distributed in large samples" even if the observations are not jointly normally distributed. However, the phrase "large samples" is quite vague, and could refer to either  $N$ , the number of days, or  $m$ , the number of sites, or both, being large. If  $m$  is large but  $N$  is not, then there is no reason to think that  $\hat{\mu}_k$  will be approximately normally distributed, despite the fact that the "sample size",  $mN$ , is large. A second example is in Section 4.3, where a reference to Mardia and Marshall (1984) is made to support a conjecture that the maximum likelihood estimates of the parameters in (4.1) will have the usual asymptotic normal distribution. The result of Mardia and Marshall (1984) requires that the size of the observation region grows as the number of observation sites grow. In the present problem, the observation region, Ireland, is unlikely to grow to satisfy someone's theorem. Stein (1987, 1988) considers inferences for spatial processes based on an increasing number of observations in a fixed region. In any case, the model given by (4.1) can be thought of as a multiple time series model, and I would guess that the parameter estimates are in fact asymptotically normal as  $N$  increases.

Another problem I would like to raise is making inferences about a spatial correlation function over distances less than the shortest distance between any two observation sites. Beyond the restriction that correlation functions be positive definite, there is no logical constraint on the form of the correlation function over these distances. In particular, Figure 3 shows some evidence of the correlation function flattening out over shorter distances, in which case, the authors' estimate of the nugget effect would tend to be too small. While misspecification of the form of the correlation function over these distances would not effect the results of the authors' cross-validation studies, it would effect inferences at a new site which was very close to one of the existing sites.

Stein, M.L. (1987) Minimum norm quadratic estimation of spatial variograms. *J. Amer. Statist. Assoc.* 82, 765-772.

Stein, M.L. (1988) Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* 16, 55-63.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

-----

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)	1. Statistics Department Stanford University Sequoia Hall Stanford, CA 94305-4065 U.S.A.
2. Name (incl. title)	Professor Paul Switzer
3. Affiliation (as you wish it to appear on your printed contribution)	Stanford University

-----

Text of Contribution (Double spaced)

(paper read at the RSS meeting 25 May, 1988)

When estimating the value of spatial processes at unobserved sites from data at observed sites the specification of the spatial correlation structure can be of major importance. The approach used by Haslett and Raftery is to approximate the contemporaneous correlation  $r_{ij}$  between any two sites  $i, j$  by a fitted exponential function of the corresponding inter-site distance. Such a smoothing and parameterization of spatial correlation has two immediate advantages-- it allows reasonable estimation of the spatial correlation structure when there is little or no time replication and it gives the needed estimates of correlations between observed and unobserved sites.

However, when there is substantial time replication, as there appears to be with these Irish wind data, then the  $r_{ij}$  will be well determined for every pair of existing sites. These well determined inter-site correlations will typically not all agree with any simple parametric function of inter-site distance. Indeed, it is noted in Figure 3 that correlations involving the Rosslare site fit poorly to the assumed exponential correlation model, and this station is removed from subsequent analyses. If there might be potential sites of interest nearby, the removal of Rosslare from the analyses could constitute an important waste of available data.

Considering the substantial amount of time replication available from these data, it would seem preferable to avoid parameterizing the correlations between existing twelve sites. In the absence of a purely distance-dependent correlation model one needs an alternative method to estimate correlations between the data sites and potential unobserved sites. A suggestion for such a program has been made by Switzer (1988). The suggestion uses both the fitted parametric correlation model and the directly estimated correlations between data sites for this purpose.

Specifically, let  $R$  and  $\tilde{R}$  respectively be  $12 \times 12$  correlation matrices between pairs of sites, the first estimated directly from each pair of observed time series and the latter obtained from the fitted exponential correlation model, say. Further, let  $R_k^*$  and  $\tilde{R}_k$  respectively denote  $12 \times 1$  correlation vectors between the putative site  $k$  and each of the 12 data sites, the first given by the expression below and the latter obtained from the exponential correlation model. As the putative site  $k$  approaches an observed site  $i$ , then the proposed  $R_k^*$  vector coincides with the  $i$ -th column of the directly estimated correlation matrix  $R$ . Other properties of the proposal are described in the above-cited report. The proposal is

$$R_k^* = [ R \tilde{R}^{-1} ] \tilde{R}_k .$$

**Winson Taam and Brian S. Yandell** (University of Wisconsin-Madison): It is a pleasure to congratulate the authors for an interesting and thought provoking investigation on the problem of modelling processes in space and time. We wish to comment on a few aspects of the model structure and computational efficiency.

The authors have chosen to use an exponential structure to model the spatial dependence among these unequally spaced weather stations. Haslett and Raftery also indicated that another approach would be to collect data on a denser grid of locations. Given an equally space rectangular lattice, the space-time model will be essentially the same as the one discussed by the authors except that the spatial structure is being modelled by a specific class of spatial models in place of the exponential correlation structure. In particular, the spatial correlation can have a spatial ARMA structure defined in Besag (1972) or Tjostheim (1978). One needs to estimate the covariance matrix for the likelihood estimation. Because of the regular grid structure, one can use a torus to approximate the covariance  $\mathbf{R}$ . Taam (1988) has indicated the approximation rate for that spectral approximation. The advantages of this approach include modelling the local spatial dependency, simplifying the computation of likelihood estimates for the spatial portion and representing the spatial structure in spectral terms. This last feature can answer the question Mr. Haslett and Mr. Raftery asked at the end of section 4.3. This approach is one way to handle the boundary problem when a likelihood estimation is used. The fractional differencing may still be used in the temporal part of the model because we have proposed an alternative way to model the spatial part of the model if the data were collected from a rectangular lattice.

It seems that one could relax the parametric nature of the Haslett-Raftery model by setting the problem in a Bayesian context of multivariate smoothing splines (Wahba, 1985; Wahba, 1983). Consider the model

$$X_{it} = f_i(t) + \epsilon_{it}$$

with  $\epsilon_{it}$  iid normal with variance  $\sigma_\epsilon^2$  and  $f_i(t)$  having a multivariate normal distribution in time and space. The covariance for  $f_i(t)$  could be (1) completely general (symmetric nonnegative definite, but no further structure); (2) a Kronecker product of a spatial and a temporal covariance; or (3) a Kronecker sum of a spatial and a temporal covariance. Case (2) includes the model considered by Haslett and Raftery as a special case. Model

(3) is much simpler, with correlated means but no cross-correlation over time. This hierarchy of models provides a framework for testing model adequacy, and avoids the parametric assumptions made in this interesting paper. This nonparametric approach may be viewed as an exploratory method to identify a model, or as a means to confirm the adequacy of a parametric model (Cox et al., 1988). The computational cost is likely to be considerable. Bates et al. (1987) provided a general algorithm for multivariate smoothing splines and indicated that without paying special attention to the design, computation becomes prohibitive on a VAX with over 400 data points. One can use the ideas in Yandell (1988) on block diagonalization to modify one dimensional spline code (Hutchinson, 1984; Reinsch, 1967) to compute estimates for (3) quickly. This same idea may also help reduce computation for case (2), although this has not been investigated.

Bates, D. M., Lindstrom, M. J., Wahba, G. and Yandell, B. S. (1987) GCVPACK - Routines for Generalized Cross Validation. *Comm. Statist. B- Simul. Comput.*, **16**, 263-297. (Algorithms Section)

Besag, J. E. (1972) Nearest-neighbor systems and the auto-logistic binary data. *J. Roy. Statist. Soc. Ser. B*, **34**, 75-83.

Cox, D. D., Koh, E., Wahba, G. and Yandell, B. S. (1988) Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.*, **16**, 113-119.

Hutchinson, M. F. (1984) A summary of some surface fitting and contouring programs for noisy data. Technical Report #ACT84/6, Div. Math. and Stat., CSIRO.

Reinsch, C. H. (1967) Smoothing by spline functions. *Numer. Math.*, **10**, 177-183.

Taam, W. (1988) A Semi-parametric Approach to Spatially Correlated Data. unpublished Ph.D. Thesis, University of Wisconsin-Madison.

Tjostheim, Dag (1978) Statistical Spatial Series Modelling. *Adv. Appl. Probab.*, **10**,

130-154.

Wahba, G. (1985) A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Statist.*, **13**, 1378-1402.

Wahba, G. (1983) Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B*, **45**, 133-150.

Yandell, B. S. (1988) Block diagonal smoothing splines. *Statist. and Probab. Letters*, **6**, 000-000.

Professor D.M. Titterington (University of Glasgow)

and Mr. P. Jamieson (James Howden & Co. Ltd.).

We should like to comment briefly on the body of the paper and to make further remarks about an aspect of wind power referred to right at the end of Section 6.

The first comment is to continue the Rosslare saga. No matter where the port is relocated as a result of the paper and discussion (the Goons would have made much of over-land ferries to Ireland!), the Rosslare data should surely be incorporated at some stage. Figure 3 suggests that this should be feasible, using a different  $\beta$ .

The second remark is to wonder whether or not the methods of the paper can be developed to create contour maps of wind speed and/or direction. With the incorporation of the time variable, these could lead to fascinating animated films of the wind behaviour over Ireland. (This could have been of particular interest to one of us who was almost blown off the sea while sailing near Cork in 1970!)

Of more serious interest to us, however, is the problem of high winds and the associated loadings imposed on wind turbines. In view of the high cost of these machines and the length of time (about 25 years) envisaged for their period of service, it is very important to be able to predict long-term extremes of wind and to translate these into extremes of stress on the turbines. While there are adequate models for the latter from the literature on structures, the complicated statistical description of wind-speeds at even a single location precludes the availability of analytical solutions, so far as extreme wind speeds are concerned. Our investigations so far have accordingly taken the form of simulation exercises.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2BH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

---

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)	1. <u>Institute of Mathematics,</u> <u>University of Kent at Canterbury,</u> <u>Canterbury, Kent, CT2 7NF.</u>
2. Name (incl. title)	<u>Professor H. Tong.</u>
3. Affiliation (as you wish it to appear on your printed contribution)	<u>University of Kent at Canterbury.</u>

---

Text of Contribution (Double spaced)

Spatial time series models are as important as partial differential equation models in the hard sciences. As a one-sided man, I admire our dexterous colleagues.

(i) In tonight's approach, spatial dependence is modelled in (4.1) via the  $\epsilon_{it}$ 's. This is similar in spirit to the 'diagonal' approach of Chan and Wallis (1978) in multiple time series. In the present context,  $E[X_{it} | X^{t-1}]$  does not depend on  $X_j^{t-1}$ ,  $j \neq i$ . Am I right in suspecting that this could be a serious constraint? Without non-parametric regression estimates of these available, I could not tell if substantial information might not be lost due to the assumption. I suspect it would if the new station is close to one of the synoptic stations, and if the time scale is short.  $E[X_{it} | X_{js}]$  could well be non-linear too!

(ii) It always strikes me that it is rather artificial and time consuming to model long-range memory by fractional differencing. I would personally feel that a Markovian model such as a non-linear autoregression (NLAR) would be a much more natural way to go about it. The snag is that it does not seem so easy to identify a suitable NLAR. Last summer H. Kunsch, D. Tjøstheim and myself were playing around with NLAR models of the form below with that objective in mind:

$$X_t = X_{t-1} + \alpha I(X_{t-1} \leq 0) - \beta I(X_{t-1} > 0) + \epsilon_t$$
$$(\alpha > 0, \beta > 0),$$
 where  $I$  is an indicator function. (Note that the model is a random walk if  $\alpha = \beta = 0$ ). It is ergodic. The hope is that it is neither geometric ergodic nor mixing! Unfortunately we ran out of time and we had to return to our respective spatial co-ordinates.

(iii) In addition to Fig. 5, it would be informative to have periodograms before the AR(9) filter.

#### References

Chan, W.Y.T. and Wallis, K.F. (1978) "Multiple time series modelling: Another look at the mink-muskrat interaction". *Appl. Statist.*, 27, 168-175.

ROYAL STATISTICAL SOCIETY  
25 Enford Street London W1H 2RH

Contribution to the discussion by Haslett & Raftery on 25 May 1988. The contribution intended for publication must be under 400 words and reach us by 6 June. It should be submitted on this sheet, in double-spaced typing. The above deadline is important (i) for the author(s) of the read paper who will consider all the contributions and compose a reply, in a limited time; and (ii) for the Journal's production. Please send your contribution to the Executive Secretary.

1. Full address where you wish to receive proofs of your contribution for checking (ie where they will reach you, approx. 3 months after the date of the meeting)

1. B.P. Exploration Co. Ltd,  
Britannic House, Moor Lane,  
London EC2Y 9BU

2. Name (incl. title)

Dr. Andrew Walden

3. Affiliation (as you wish it to appear on your printed contribution)

B.P. Exploration Co. Ltd.

Text of Contribution (Double spaced)

Haslett and Raftery are to be congratulated for clearly and appealingly applying a variety of statistical techniques — some well established, others less so — to an important practical problem.

The long-memory temporal dependence raises some interesting questions. The authors acknowledge the main problem in recognizing long-memory dependence, viz it is difficult or impossible in practice to distinguish between spectral shape caused by truncating the autocovariance function of a long-memory process (through the use of a finite sample) from spectral shape arising from a process which does not satisfy the long-memory model. Several of the spectra of fig. 5 show decay rates of 12dB/octave (i.e.,  $f^{-4}$ ) at a frequency as low as 0.0005. By restricting  $d$  to  $0 \leq d \leq 0.5$ , the authors implicitly restrict frequency decay rates to be no greater than  $f^{-1}$  at such low frequencies. Do the authors feel that the problem referred to above is sufficient explanation of this discrepancy? Did they consider spectral approaches to the estimation of  $d$  such as that of Janacek (1982) ?

It is interesting to consider physical mechanisms for red-noise spectra similar to those seen in fig. 5. An ensemble of purely random processes, each with an autocovariance of the form  $e^{-|r|/\tau_0}$  and its own correlation time  $\tau_0$  can generate red-noise spectra with differing decay rates in different frequency ranges depending on the distribution of  $\tau_0$ . This has been used to model the river level at the mouth of the Nile (Montroll and Shlesinger, 1982) for which the predominant decay is  $f^{-1}$ . Mechanisms for higher decay rates are discussed in Halford (1968).

#### REFERENCES

Halford, D.(1968) A general mechanical model for  $|f|^\alpha$  spectral density random noise with special reference to flicker noise  $1/|f|$ . *Proc. IEEE*, **56**, 251-258.

Janacek, G.J.(1982) Determining the degree of differencing for time series via the log spectrum. *J. Time Series Analysis*, **3**, 177-183.

Montroll, E.W., and Shlesinger, M.F.(1982) On  $1/f$  noise and other distributions with long tails. *Proc. Natl. Acad. Sci. USA*, **79**, 3380-3383.